Allosteric Regulation of V(D)J Recombination by the Active Chromatin Mark H3K4me3

by
John Thomas Bettridge

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

September, 2017

Abstract

V(D)J recombination is initiated by the RAG recombinase; a heterotetrameric complex consisting of two RAG-1 and two RAG-2 subunits. The susceptibility of gene segments to cleavage by RAG is associated with histone modifications characteristic of active chromatin, including pan histone acetylation, and trimethylation of histone H3 at lysine 4 (H3K4me3). Binding of H3K4me3 by a plant homeodomain (PHD) in RAG-2 stimulates both DNA substrate binding and catalysis of the RAG-1 subunit. This has suggested an allosteric mechanism in which information regarding occupancy of the RAG-2 PHD is transmitted to RAG-1, thereby altering RAG-1's activity. To determine whether the conformational distribution of RAG is altered by H3K4me3, we mapped changes in solvent accessibility of cysteine thiols by differential isotopic chemical footprinting using tandem liquid chromatography mass spectrometry. Binding of H3K4me3 to the RAG-2 PHD induces conformational changes in RAG-1 within the DNA-binding domains and in the ZnH2 domain, which acts as a scaffold for the catalytic center. These conformational changes are dependent on an intact PHD finger: mutation of tryptophan W453 in RAG-2 to alanine disrupts the RAG-2 PHD-H3K4me3 interacton, and renders RAG-2 unable to stimulate conformational changes in RAG-1. Thus, engagement of H3K4me3 by an intact RAG-2 PHD is associated with dynamic conformational changes in RAG-1, consistent with allosteric control by active chromatin. To reveal the interface through which RAG-2 transmits this allosteric signal to RAG-1, we generate chimeric murine RAG-2 proteins bearing plant homeodomains derived from distantly related shark RAG-2 PHD fingers. One of these chimeric RAG-2 proteins, bearing a PHD finger derived from C. *punctatum*, exhibits a loss of function similar to W453A, but retained the ability to bind H3K4me3.

Selective reversion of amino acid residues 417 through 433 in the C. *punctatum* PHD finger back to murine sequence was sufficient to rescue the W453A-like phenotype, indicating the allosteric transmission interface lies within this region of RAG-2. These data support a model where RAG-2 engagement of H3K4me3 produces conformational changes within the surface of the RAG-2 PHD finger, which in turn directly stimulate the activity of RAG-1, "priming" the RAG-1 subunit for interactions with DNA.

Ph.D. Dissertation referees for John Thomas Bettridge

Stephen Desiderio, M.D., Ph.D.

Director, Institute for Basic Biomedical Sciences

Professor, Molecular Biology and Genetics

Johns Hopkins University School of Medicine

(Thesis advisor)

Ranjan Sen, Ph.D.

Chief, Laboratory of Cellular and Molecular Biology

National Institute on Aging

(Reader)

Acknowledgements

Dr. Stephen Desiderio has been my single most formative mentor as a young scientist to date. His encouragement, commitment to my project, and guidance throughout my thesis guided me to success, and helped me develop new sets of skills. His mentorship style of high expectations, but giving me independence to pursue almost any project brought both significant insights in some instances, and in other cases, important lessons in careful experimental design. Through my entire education he has been a supportive mentor, insightful investigator, and exceptional teacher. My training under his tutelage has been nothing short of exceptional.

I also want to extend my most sincere gratitude to my wife, Kelsey Bettridge. She is a wonderful life companion, constant supporter of me and my work, and an accomplished, brilliant biophysicist. Without her I doubt I would have completed my doctorate; Kelsey I am overjoyed every day to be your husband.

I would like to extend my gratitude to my labmates, Alyssa Ward, Meiling May, Nick Dordia, and Wenyan Lu. You have been wonderful labmates, peers, collaborators, and friends. I will greatly miss the time I spent discussing new experiments and data with you all, and the incredible lab environment you all helped create. I'd also like to acknowledge Chao Lu, who taught me biochemistry and was also a great part of our lab, I wish him all the success in his endeavors as a principle investigator. I would also like to thank my collaborators, Dr. Akhilesh Pandey and Dr. Chan-Hyun Na, who were essential in publishing my 2017 PNAS paper. I am also indebted to my thesis committee: Dr. Joel Pomerantz, Dr. Ranjan Sen, and Dr. Vincent Hilser, who encouraged me to be the best

scientist I could be, and provided crucial insights into my research; you helped me become a better scientist and critical thinker.

I want to thank my mother and father, John and Beth Bettridge, for their undying support, through my doctorate and throughout my life. I appreciate you all patiently listening to my scientific ramblings at Thanksgiving and Christmas, and will always remember our wonderful trips to New York and Florida during my thesis. The kindness you have shown me is unparalleled, and I am so grateful for everything you've done for me.

Lastly I wanted to thank my BCMB friends; the close classmates and friends who supported me throughout my Ph.D. and helped make the difficult road a little easier: Risa Burr, Catherine Gilbert, Laurel Oldach, Anthony Schuller, Jarrett Smith, Stefanie Tan, Jason Westerbeck, and Mark Zbinden. Thank you all for being wonderful peers and friends: we are a unique group, and I hope our friendships endure long into the future.

Table of Contents

List of Figures

Introduction

V(D)J recombination and the immune system

The human body is challenged by thousands of viruses, bacteria, and other potential pathogens every day. To prevent and eliminate infections, the immune system must have a means of recognizing these foreign invaders. Recognition of foreign agents is achieved through the production of antigen receptors, such as immunoglobulins. These antigen receptors can recognize almost any foreign antigen the body encounters, stimulating the immune system to seek out and destroy these potential pathogens.

In immature lymphocytes, antigen receptor genes are encoded as discreet, non-contiguous gene segments, known as the Variable (V), Diversity (D), and Joining (J) segments. To form a functional antigen receptor gene, these gene segments must be assembled through a specialized form of DNA recombination called V(D)J recombination. V(D)J recombination is initiated by the recombination activating genes, RAG-1 and RAG-2, which form a heterotetrameric recombinase complex known as the RAG recombinase (RAG). RAG induces double strand breaks at recombination signal sequences (RSSs) which flank the V, D, and J segments (Schatz and Swanson 2011). These double strand breaks are then resolved by the non-homologous end joining (NHEJ) pathway, which join the gene segments together to form a functional antigen receptor. Each V, D, and J gene segment family has multiple members, all of which can be joined with members of the other families to produce an extraordinary number (3×10^{11}) of V-D-J combinations. This immense diversity is what allows the immune system to generate antigen receptors that recognize nearly any epitope.

The two ultimate protein products of V(D)J recombination assembled antigen receptors, are antibodies and T cell receptors. Antibodies composed of two “heavy chain”

polypeptides, and two "light chain" polypeptides, which are encoded at the immunoglobulin heavy chain locus (IgH) and the immunoglobulin κ or λ loci respectively. These four peptides are assembled in a "Y" shaped structure, which is stabilized by disulfide bonds that connect the heavy and light chains together. T-cell receptors are encoded by the TCR α and β, or γ and δ loci; polypeptides generated from these loci form a heterodimer of either α and β peptides, or γ and δ peptides. These heterodimers exhibit disulfide linkages similar to antibodies, and are expressed as membrane anchored proteins on the surface of T-cells.

Because RAG is an endonuclease, its activity must be carefully controlled to prevent the generation of off-target DNA double strand breaks, which can lead to oncogenic genomic lesions. There are multiple levels of control imparted on RAG by the cells in which it is expressed. For example, V(D)J recombination is cell lineage specific: in developing B cells, only immunoglobulin genes are rearranged, while the T Cell Receptor (TCR) remains silent, whereas in immature T Cells, the TCR is recombined while the immunoglobulin genes remain unrearranged (Schatz and Ji 2011). Despite the potential oncogenicity of RAG, its activity is essential to healthy immune function. Inactivating mutations of either subunits of RAG result in severe combined immunodeficiency (SCID); that is the inability to generate B or T cells. Patients with this disorder are highly susceptible to infections and are unable to combat immune challenges.

The recombination activating genes RAG-1 and RAG-2

The RAG recombinase is a heterotetramer of the RAG-1 and RAG-2 proteins, which are 1040 and 527 amino acids long, respectively. Each protein is delineated by a "core" domain, and one or more "non-core" domains: the core domains are the minimal protein fragments which will support DNA cleavage *in vitro*, while the non-core domains contain regulatory elements which regulate the activity of the RAG protein complex.

RAG-1

The RAG-1 core domain is composed of seven distinct domains: an N-terminal nonamer binding domain (NBD) is connected to a Dimerization and DNA Binding Domain (DDBD) by a flexible linker (Fig1A). The DDBD in turn connects to a Pre-RNASE H (PRER) and RNASE H fold (RNH), that comprise the catatlytic site of RAG-1 where DNA cleavage occurs (Kim 2015). The two RAG-1 active sites within the tetramer are spaced ~45 angstroms apart in three-dimensional space. Beyond the RNASE H catalytic fold are the ZnC_2 and ZnH_2 domains, which form a zinc binding site, an interface for binding to RAG-2, and a portion of the DNA binding domain (Kim 2015). The uncrystallized N-terminal non-core domain of RAG-1 contains a RING domain of unclear function, which is dispensable for recombination (Yurchenko 2003). However, the RING domain has demonstrated autoubiquitilation activity, which may regulate the activity of RAG-1, but further study is needed to determine its role, if any, in V(D)J recombination.

RAG-2

In RAG-2, the core domain comprising residues 1-352 is structured into a six-bladed β propeller, composed of six individual kelch domain repeats (Kim 2015). These kelch domains make direct contact with RAG-1, and are thought to contribute to DNA binding (Kim 2015, Ru 2015). Mutations within core RAG-2 have been identified that impair recombination activity (Qui 2001): these mutations either disrupt contacts between RAG-1 and RAG-2, or ablate RAG-2 contacts with DNA within the active site (Kim 2015). The non-core region of RAG-2, comprising residues 352-521, is dispensable for RAG mediated DNA cleavage in vitro (Silver 1993, Liang 2002). However, removal of this non-core region disrupts endogenous recombination and dysregulates RAG expression in lymphocytes (Liang 2002, Mijuskovic 2015). Several important regulatory domains have been identified in RAG-2, including an autoinhibitory domain (residues 352-405), a plant homeodomain (PHD) finger (residues 405-487), and a cell cycle regulatory site (T490) (Fig 1B). The autoinhibitory domain, described in detail in other publications from our lab (Lu 2007, Ward unpublished data) is thought to be an unstructured loop that is highly negatively charged due to high glutamate and aspartate composition. Mutation of the acidic residues in this domain rendered RAG "hyperactive", enhancing both the catalytic rate and binding affinity of RAG (Lu 2015). In addition to the autoinhibitory domain, RAG-2 possesses a CDK phosphorylation site, threonine T490. T490, when phosphorylated by CDK2, targets RAG-2 for degradation through the Skp2-SCF ubquitination pathway (Li 1996). CDK2 is activated and phosphorylates RAG-2 at the G1-S transition, signaling its degradation and thereby restricting expression of RAG-2 to the G1 cell phase (Zhang 2011). This is to ensure genomic stability: in

G1/G0, the NHEJ repair pathway predominates, so RAG mediated double strand breaks are repaired by NHEJ. However, if RAG-2 were allow to persist into S phase, where homologous recombination predominates, RAG mediated breaks could be repaired by HR, potentially resulting in chromosomal translocations (Papaemmanuil 2014). Indeed, RAG-2 proteins bearing a T490A mutation persist into S phase, and mice bearing this gene exhibit increased chromosomal abberations and develop leukemias on a P53 $^{-/-}$ background (Zhang 2011).

The PHD finger of RAG-2 is the only domain of the non-core region to have been crystallized to date: this domain's architecture is similar to that of the YNG-2 and ING PHD fingers: a zinc finger domain with a pocket for binding the trimethylated ammonium of histone H3 lysine 4. Within this pocket, three residues make critical contacts with the trimethylammonium, creating a hydrophobic channel that stabilizes the interaction between RAG-2 and H3K4me3: Y415, W453, and M443 (Ramon-Maiques 2007, Shimazaki and Lieber 2014). Mutation of any of these three residues has been shown to be highly detrimental to recombination, rendering the RAG recombinase nearly inactive (Lu 2007). In particular, mutation of tryptophan W453 has been known to cause SCID syndrome in patients, and exhibits a near complete loss of function in vivo (Ramón-Maiques 2007; Matthews 2007; Liu 2007). Evidence suggests that this is due to several effects, including the disruption of Pi stacking contacts between the tryptophan and the methyllysine, and destabilization of the structure of the PHD finger through the elimination of a bulky hydrophobic residue. Paradoxically removal of the entire non-core domain of RAG-2 is less detrimental than mutation of W453 in the RAG-2 PHD finger, suggesting that the PHD finger of RAG is more than merely a binding site, and may

exhibit unique functions (Lu 2007). This phenomenon is the subject of this thesis, and is discussed in detail in the following chapters.

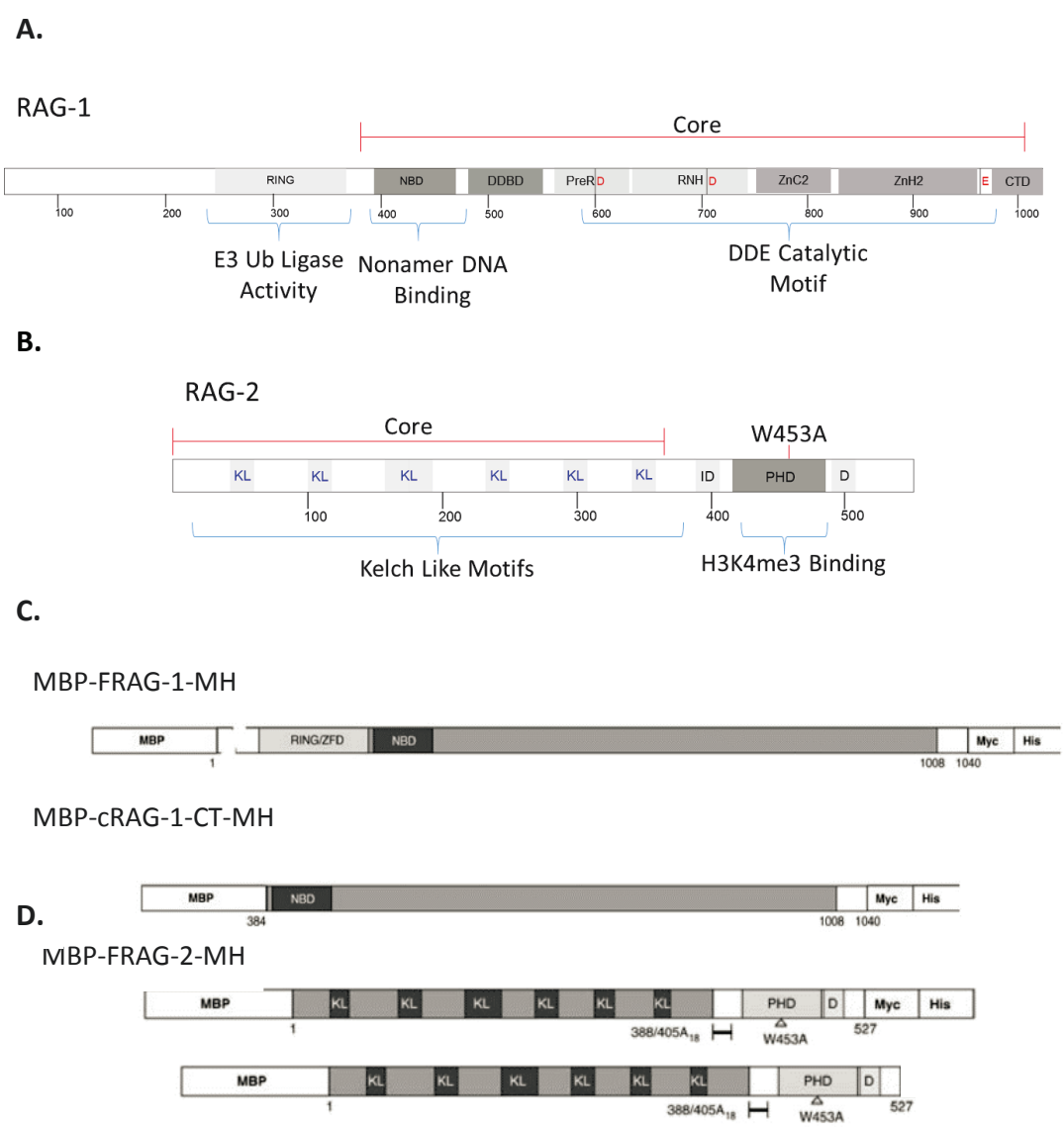


Figure 1.1 Diagrams of the RAG proteins A. Detailed diagrams of the RAG-1 and RAG-2 proteins, with the "core" regions defined in each. RAG-1: the RING-type zinc-finger domain (RING/ZFD) and the nonamer binding domain (NBD), and additional domains (defined in text) are indicated. Red letters, catalytic residues. RAG-2: . Kelch-like domains (KL), inhibitory domain (ID) the PHD finger (PHD), and the degradation signal (D) are indicated. The critical W453A mutation, which ablates H3K4me3 binding, is shown. B. Diagrams of wild-type RAG-1 (top), full-length RAG-1-MH (flR1-MH, middle), and core RAG-1ct-MH (cR1ct-MH, bottom). Amino acid residues at domain boundaries are numbered. The core is designated in dark gray. MBP, Myc, and His denote the maltose binding protein, c-myc epitope, and polyhistidine tags, respectively. C, RAG-2 constructs for in vitro assays. Upper: MBP fusion wild-type RAG-2. Low: full-length RAG-2 fusion protein with TEV cleavage site. Amino acid residues at domain boundaries are numbered MBP, Myc, and His are as defined in (D). Positions of mutations are marked below.

Mechanism of V(D)J recombination and the 12/23 rule

V(D)J recombination is initiated by the binding of RAG-1 and RAG-2 to a 12-RSS and a 23-RSS. Upon binding to the RSSs, the RAG proteins are brought together in 3-dimensional space to form the pre-synaptic complex, which is a heterotetramer comprising a stoichiometry of $(RAG\text{-}1)_2$ and $(RAG\text{-}2)_2$ bound to both a 12 and 23 RSS (Schatz and Swanson 2001). Only complexes containing RSSs with different spacer lengths undergo recombination, thereby restricting V(D)J recombination and preventing segments of the same gene family from joining (e.g. Jh-Jh joining), which is referred to as the 12/23 rule . This restriction is structural in nature: the RAG heterotetramer exhibits an asymmetrical "tilt" in the nonamer binding domains of the RAG-1 subunits (Kim 2015, Ru 2015). This asymmetry disallows concomitant binding of two similar RSSs, i.e. two 12 RSSs: the distance on the longer side of the tetramer is simply too far to accommodate bind a 12 RSS (Kim 2015). After synapsis, RAG mediated DNA cleavage step occurs in two distinct, biochemically separable steps: first, RAG-1 nicks both DNA segments on the leading strand at the junction between the RSS and the antigen receptor gene segment, generating two 3'-OH groups. Next, RAG promotes a transesterification reaction, in which each newly exposed 3'-OH group attacks the complementary DNA strand, generating two blunt 5'-phosphorylated signal ends, and two coding ends terminating in hairpins. After hairpins are opened by the endonuclease artemis, signal ends and coding ends are then joined by the NHEJ pathway to produce signal and coding joints, and a functional antigen receptor gene (Schatz and Swanson 2011).

Regulation of the RAG recombinase

Despite the benefits of V(D)J recombination being able to generate antigen receptors that can recognize nearly any epitope, the nuclease activity of the RAG recombinase must be carefully regulated. Improper, mistimed, and mistargeted RAG activity can lead to DNA double stranded breaks that are potentially oncogenic and destabilizing to the genome (Zhang 2011, Papaemmanuil 2014, Mijuskovic 2015). Therefore V(D)J recombination is restricted both physically, temporally, and in a lineage specific fashion. RAG-1 and RAG-2 gene expression is restricted transcriptionally to early lymphocyte development, preventing ongoing recombination through the life of the lymphocyte. V(D)J recombination is restricted to the G1 cell cycle phase by the periodic destruction of RAG-2 through phosphorylation of T490 at the G1-S transition by CDK2, which targets the protein for ubiquitination and degradation by the proteasome (Li 1996, Zhang 2011). Restriction of V(D)J recombination to G1 ensures that RAG-mediated DNA breaks are repaired by NHEJ, and not by the homologous recombination (HR) pathway, which can cause chromosomal translocations if allowed to repair V(D)J recombination initiated breaks (Zhang 2011).

Transcription and epigenetic control of the RAG recombinase

Within an antigen receptor gene locus, not all V, D, and J gene segments undergo V(D)J recombination with equal efficiency. While sequence variation within the recombination signal sequences (RSSs) can account for some of this variability, RSS usage alone cannot account for the differences observed during lymphoid development. More recently it has been appreciated that those segments which lie more proximal to transcriptional promoters, such as a DQ52 in the IgH locus, exhibit increased usage in

recombination compared to its more distal family members such as DH6 (Schatz and Ji 2011, Gigi 2014). Ultimately, it was discovered that chromatin modifications associated with these transcriptional changes are largely responsible for activation or suppression of the V(D)J recombination. A detailed discussion of these mechanisms is outline in the following section.

V(D)J recombination of an antigen receptor gene is invariably preceded by transcriptional activation of the locus. This germline transcription, termed "sterile transcription" because it produces mRNA that does not code for protein, initiates from promoters embedded within antigen receptor loci. Antigen receptor loci undergo dramatic alterations in chromatin structure prior to recombination, transitioning from a tightly packaged, Dnase resistant state to a Dnase sensitive form (Mills 2003; Yancopoulos 2012). This change is correlated with active transcription and the production of germline transcripts at these genes. Initiation of germline transcription promotes the establishment of euchromatin marks, such as histone acetylation and DNA hypomethylation (Schatz and Ji 2011). These observations prompted the accessibility hypothesis, which postulates that RAG is actively excluded from antigen receptor loci until a developmental switch alters these loci by initiating germline transcription through the gene segments to be recombined (Cobb 2006). This active transcription changes the chromatin from an inaccessible to accessible form, allowing DNA binding elements including the RAG proteins to interact.

Detailed investigations of the Ig and TCR loci revealed that germline transcription stems from multiple promoters located proximal to the V, D, and J gene segment clusters. These promoters are activated in a developmental and lineage-dependent fashion prior to

recombination: sterile transcription is restricted primarily to immunoglobulin (IgH, Igκ, Igλ) loci in developing B cells, and at TCRα, TCRβ, TCRγ, and TCRδ in thymocyte precursors (Van Ness 1981; Yancopoulos 1985).

Transcriptional regulation of recombination in B cells

The IgH locus can be divided into 3 general regions: the most 5' region containing the Vh gene segments, a central region which contains the Dh and Jh gene segements, as well as the principal IgH enhancer Eμ and the intergenic control region (IGCR1), and the 3' region which encodes the various constant region isotypes (Subrahmanyam and Sen 2010). Within the immunoglobulin heavy chain (IgH) locus, the Vh gene segments are flanked by both sense and anti-sense promoters, which are inactive during Dh to Jh joining, but are activated prior to Vh-DJh joining (Bolland 2004, Bolland 2016). The earliest detectable germline transcription at IgH stems from the D segment promoter pDQ52, located between the D and J segments, and from the Eμ enhancer (Schlissel 1991). However, pDQ52 is dispensable for V(D)J recombination, although Eμ is required for efficient recombination (Afshar 2006).

Transcription at the IgH locus is governed by several enhancers that interact with both the Vh and the Dh and Jh promoters. The primary IgH enhancer, Eμ, lies 3' of Jh gene segments and 5' of the Cμ exons (Cobb 2006). Deletion of Eμ ablates germline transcription over Dh and Jh segments, and blocks V-DJ joining: D-J still occurs, albeit at a reduced frequency. These results indicate that additional and/or redundant mechanisms might regulate or permit the first step in recombination (D-J joining) even in the absence of the enhancer (Afshar 2006, Perlot 2005, Cobb 2006). One of these redundant elements may be cμ, which has exhibited both promoter and enhancer activity (Jung 2006).

In addition to promoter and enhancer elements, the IgH locus is subject to control by an intergenic control region, IGCR1. IGCR1 insulates Vh segments that lie proximal to Dh promoters from crossover transcription. They also function, through an unconfirmed mechanism, to normalize Vh usage by discouraging proximal Vh use and enhancing distal Vh recombination. It is suspected that CTCF is responsible for this activity by binding to IGCR1 and creating chromatin loops that bring distal Vh segments into proximity with the Dh region, promoting distal Vh to DJh joining (Guo 2011).

Transcriptional regulation in developing thymocytes

Immature thymocytes initiate V(D)J recombination early in their development, and target TRCβ for recombination first. Within the TCRβ locus, two promoters and a single enhancer govern transcription over the DβJβ clusters. The first promoter, PDβ1 is positioned 5' of the Dβ1 and Jβ1 gene segment families, and the second, PDβ2 lies 3' of the DJβ1 cluster, although its precise location has not been verified (Doty 1999, Sikes 2011). Additional promoters are interspersed within the Vβ clusters and control local transcription of the V β families. The single enhancer (Eβ) is located near the 3' end of the locus, and promotes transcription across both the Dβ1 and Dβ2 clusters (Whitehurst 2000). Deletion of PDβ1 disrupts V(D)J recombination of Dβ1-Jβ1 segments, but does not affect Dβ2-Jβ2 recombination or subsequenct Vβ-DJβ joining (Whitehurst 2000). Mice harboring an Eβ deletion are unable to intiate recombination. In contrast, deletion of Eβ, which is activated at the earliest thymocyte precursor stage, oblates transcription across both clusters locus (McDougall 1988). Concurrent with this loss of transcription, heterochromatin is established across TCRβ DβJβ clusters, inhibiting Dβ-Jβ

recombination. Together, these results suggest that PDβ1 and PDβ2 independently interact with the Eβ enhancer

Transcription factors involved in V(D)J recombination

Only a handful of transcription factors have been found to participate in transcription of the of antigen receptor genes, and have most have been identified in B cells. E2A and EBF, which are alternative splice variants of the same gene, have been found to bind to the IgH Eμ enhancer. Deletion of either E2A or EBF causes a block in D-J joining, and inhibits transcription across the DJ clusters (Bain 1999). The paired box factor, Pax5, is a B cell specific transcription factor that is expressed early in B cell development, and has known binding sites within the IgH locus at the Eμ enhancer. Expression of Pax5 promotes proximal Vh-DJh rearrangement, and interestingly, this function appears to be independent of germline transcription (Fuxa 2004). Although deletion of TFs severely impairs lymphocyte development, deletion of their promoters is less detrimental indicating that these factors might play an additional role besides promoting germline transcription.

Large scale changes of chromatin in response to transcription

Transcriptional activation and repression of these genes is associated with post-translational modifications of histones: such as acetylation, methylation. Similarly, during transcriptional silencing of these loci, the primary modification deposited on the histones is methylation. Attempts to uncouple germline transcription from V(D)J recombination have not been successful, indicating that transcription and its associated processes are essential for recombination. However, transcription of an antigen receptor locus alone is

not sufficient to drive recombination. For example, in PAX5 deficient cells, some Vh segments are still transcribed, but never undergo recombination (Fuxa 2005). From these data it appears that the role of transcription is more than just generating accessible antigen receptor loci by altering the local chromatin, but rather has profound effects on DNA looping and the topology of antigen receptor loci (Chaumeil 2013). Indeed, recent findings suggest that the transcription factors recruited during germline transcription are of equal or greater importance to successful recombination than transcription alone. Chromosomal looping effected by factors such as CTCF and PAX5, which bind the IGCR and Eμ respectively have been proposed to restrict V(D)J recombination to particular segments by trapping RAG within a CTCF-generated chromatin loop (Bolland 2016, Lin 2015, Feehley 2012). Deletion of the IGCR in the IGH locus results in increased proximal V-DJ joining, and a general increase in the amount of overall recombination in IGCR deficient cells, which could be explained by the inclusion of proximal Vh segments within a chromosomal loop with the Dh and Jh segments when IGCR is deleted (Hu 2015, Guo 2011).

In addition to alterations in chromatin looping and structure, transcription may also alter DNA nucleosome interactions directly. Bevington et al. report that active transcription within the Vκ locus may transiently displace the H2A/H2B heterodimer from nucleosomes, resulting in altered nucleosome DNA contacts that allow RAG to bind and cleave at RSSs (Bevington 2013). Nucleosome remodeling factors such as SWI/SNF have also been proposed to regulate germline transcription and locus accessibility at IgH through interactions with Eμ (Osipovich 2009). SWI/SNF has also been proposed to alter interactions of RSSs with nucleosomes, sliding the nucleosomes so that RSSs are

repositioned in a manner that allows RAG to interact directly with the RSS as free unbound DNA (Baumann 2003, Kassabov 2003, Shimazaki 2014). It is difficult to parse out when these various mechanisms are active during B-cell development, and whether they work in concert with one another, work as redundant mechanisms for accessibility, or are simply a non-specific byproduct of general transcription. Regardless, the transcription factors recruited during germline transcription appear to be the main effectors in driving V(D)J recombination, rather than the general accessibility of DNA generated by transcription.

Epigenetic modification in V(D)J recombination

Mammalian genomes are organized within the nucleus into chromatin: a DNA-protein hybrid in which winds DNA around nucleosomes, thereby packing the DNA into a much more compact form than naked, free DNA. Nucleosomes are composed of 4 histone proteins, H1, H2A/H2B, H3, H4. There are multiple isoforms of the proteins, but a discussion of their roles is beyond the scope of this thesis. A nucleosome is composed of two copies of H2A, H2B, H3, and H4: DNA is wound around this protein complex like thread on a spool. Each nucleosome winds ~146 base pairs of DNA. Nucleosomes are naturally "phased" across the genome, meaning they show a periodic distribution of ~150 base pairs (Strahl 2000). This phasing can be altered by a number of chromatin remodeling complexes including the SWI/SNF ATP dependent nucleosome remodeling complex. It has been suggested that nucleosome phasing over RSSs and antigen receptor gene segments may control RAG activity by permitting or rejecting RAG binding to the RSS, Several groups have shown that the remodeling of nucleosomes by SWI/SNF to expose the RSS as naked DNA permits RAG binding and recombination of antigen

receptor gene segments, but these data are somewhat unclear, as SWI/SNF remodeling also leads to hyperacetylation of the histones, which was shown to independently stimulate recombination (Shimazaki 2014, Roth 2000). Positioning of antigen receptor loci within the nucleus also appears to affect susceptibility to undergo rearrangement by V(D)J recombination. In maturing B cells, the IgH and Igκ loci migrate from the nuclear periphery to a more centralized position within the nucleus (Kosak 2002). The IgH locus migrates before Igκ; this migration coincides with the initiation of Dh-Jh recombination. Upon productive rearrangement of one allele, recombination of the incompletely rearranged second allele is suppressed by migration of the allele back to the nuclear periphery, and rearrangement of the light chain (Igκ) is initiated (Kosak 2002, Fuxa 2004).

In addition to large scale chromatin architecture, it appears that the transcriptionally associated histone post-translational modifications are critical in the regulation of V(D)J recombination, particularly modifications of histones H3 and H4. These modifications, and their associated transcriptional status are detailed below.

Histone H3 Lysine 9 methylation

In transcriptionally silent chromatin, e.g. DNA which is inaccessible to RNA or DNA polymerases or transcription factors, histone H3 is dimethylated at lysine 9 (H3K9me2). In B and T cells, deposition of H3K9me2 is carried about the G9a methyltransferase. In developing B-cells, H3K9me2 is initially deposited on IgH V segments (Johnson 2004). Removal of H3K9me2 in B-cells occurs in a PAX5 dependent manner, at the pro-B to pre-B transition, immediately prior to Vh-DJh joining (Johnson 2004). Deposition of H3K9me2 is sufficient to inhibit V(D)J recombination: in an elegant

experiment, Osipovich et al demonstrated that targeting of a G9a methyltransferase to a TCRβ minilocus, and subsequent H3K9me3 deposition, was sufficient to inhibit both transcription and V(D)J recombination of the minilocus (Osipovich 2004).

DNA methylation

DNA methylation is a modification of cytosine which occurs at CG dinucleotides, referred to as C_pG islands (Suzuki and Bird 2008). In uncommitted lymphoid progenitors and non-lymphoid cells, antigen receptor loci exhibit extensive C_pG methylation (Mather 1983, Storb 1983). As cells commit to the B-lineage, there is concurrent demethylation of Igκ and IgH (Mostoslavsky 1998, Selimyan 2013). A detailed analysis of B-cells at various stages of V(D)J recombination at the IgH locus demonstrated that the Dh and Jh gene segment families remain methylated prior to Dh-Jh recombination, but are found to be partially demethylated upon Dh-Jh joining (Selimyan 2013). Unsurprisingly the Eμ enhancer was found to be demethylated at all stages in these cells, likely a result of its strong enhancer function. Further, demethylation at the Dh and Jh appears to be dependent on interactions with Eμ. Mice bearing an H19 imprinting control region (ICR) inserted between Eμ and the distal Jh segments exhibited hypermethylation across both the Dh Jh clusters, and a marked decrease in germline transcription and Dh-Jh recombination (Puget 2016). Even in cells which undergo Dh-Jh joining, the DhJH recombination products remained methylated. Taken together these data suggest that the presence of DNA methylation discourages V(D)J recombination, and that temporal control of V(D)J recombination may be regulated in part by Eμ dependent demethylation of the IgH locus.

Histone H3 and H4 acetylation

Acetylation of histones H3 and H4 is a hallmark of actively transcribed chromatin. Deposition of acetylation is thought to occur concurrently with transcription, and is found across the entire transcribed locus. Histone acetylation is correlated with antigen receptor genes poised to undergo rearrangement (McMurry & Krangel 2000, Shetty & Shatz 2015). A particular acetylation mark, H3K27Ac, is suspected of direct interactions with the RAG recombinase, potentially through interactions with the RAG-1 RING domain (Maman 2016). However the functional role of this acetylation mark is unclear, particularly since histone acetylation is ubiquitous across transcribed genes, but binding of RAG-1 is specific to antigen receptors (Ji 2010).

Histone H3 Lysine 4 methylation

Like acetylation, H3K4me3 is a histone post-translational modification that is deposited in a transcription dependent manner. However, unlike histone acetylation, which shows broad distribution across actively transcribed genes, H3K4me3 is deposited in localized sites within the genome, such as the promoters and enhancers of genes (Subrahmanyam and Sen 2010). The presence of H3K4me3 is positively correlated with active V(D)J recombination, suggesting a link between recombination and this histone mark (Yancopoulos 1985). Insights into the regulation of V(D)J recombination at the epigenetic level came with the discovery that one of RAG subunits, RAG-2, contains a plant homeodomain (PHD) that specifically interacts with histone H3 when it is tri-methylated at lysine 4 (H3K4me3) (Liu 2007; Matthews 2007; Ramón-Maiques 2007). This provided the first mechanistic link between this “active chromatin” post-translational modification and V(D)J recombination. Initial experiments suggested that

the interaction between H3K4me3 and RAG-2 functioned in localization of the RAG complex (Matthews 2007). However, two lines of reasoning argue against this interpretation: first, RAG-1 is able to localize to the correct genomic location independent of RAG-2, and second, deletion of the entire RAG-2 non-core domain does not ablate V(D)J recombination, while a single point mutation within the RAG-2 PHD finger nearly eliminates recombination (Liu 2007). Unlike other PHD fingers, the RAG-2 PHD binding site for H3K4me3 contains a channel that is not filled by H3K4me3 alone. An H3K4me3 peptide containing an additional modification, symmetrical methylation of arginine 2 of histone H3, has shown to enhance RAG-2 binding to H3K4me3 by occupying this channel. H3R2me2K4Me3 exhibits enhanced affinity for the RAG-2 PHD of nearly 4 fold, suggesting that the presence of both marks might "tune" RAG-2 binding and activity *in vivo*, although ChIP experiments have shown that RAG-2 binds ubiquitously over any H3K4me3 bearing genomic region (Ji 2010, Shetty & Schatz 2015).

Regulation of RAG activity by H3K4me3

Careful enzymatic analysis revealed that binding of H3K4me3 to RAG-2 is more than just a localization signal. DNA cleavage assays utilizing a 12 and 23 RSS DNA substrate and highly purified RAG proteins demonstrated that the addition of an H3K4me3 peptide in trans actively stimulated DNA cleavage by the RAG complex (Lu 2015, Shimazaki 2012, Grundy 2010). Similarly, an H3K4me3 peptide added in trans increased the affinity of RAG for DNA substrates, and stabilized the RAG complex when bound to DNA (Liu 2007, Shimazaki 2009). Mutation of a critical tryptophan within the RAG-2 PHD finger, W453, abolishes binding of H3K4me3 to RAG-2 and also renders

RAG unstimulable by H3K4me3. These observations led us to the hypothesis that H3K4me3 is an allosteric regulator of the RAG recombinase. Evidence supporting this hypothesis is presented in the following chapters.

Chapter 1

Introduction

During lymphocyte development, the genes encoding antigen receptors are assembled from discrete gene segments by V(D)J recombination. This process is initiated by the proteins RAG-1 and RAG-2, which together bind and cleave DNA at recombination signal sequences (RSSs) that flank the participating gene segments (Gellert 2002; Schatz and Swanson, 2011). There are two classes of RSS, termed 12-RSS and 23-RSS, in which the heptamer and nonamer elements are separated by spacers of 12 bp or 23 bp, respectively. DNA cleavage by RAG involves nicking at the junction between the RSS and the coding sequence, followed by transesterification to produce a blunt, 5'-phosphorylated signal end and a coding end that terminates in a hairpin. Under physiologic conditions complete cleavage requires the pairing of a 12-RSS with a 23-RSS, so that recombination between like segments is suppressed (Schatz and Swanson, 2011).

V(D)J recombination acts in an ordered, locus-specific fashion during lymphoid development. In the B lineage, for example, the immunoglobulin heavy chain (IgH) locus is rearranged before the light chain loci and assembly of an Ig heavy chain gene proceeds by sequential D-to-Jh and Vh-to-DJh joining (Alt 1985). Upon productive rearrangement, further recombination at the IgH locus is suppressed, thereby enforcing monoallelic expression of Ig heavy chain (Jung 2006). The ability of gene segments to undergo V(D)J recombination is positively correlated with transcription at the unrearranged locus and with histone modifications characteristic of active chromatin, including hypermethylation of histone H3 at lysine 4 (Chakraborty 2007; Liu 2007; Matthews 2007; Subrahmanyam 2012).

RAG-1 and RAG-2 are 1040 and 527 amino acid residues long, respectively. Although the canonical non-core region of RAG-2, comprising residues 387 through 527, is dispensable for DNA cleavage in vitro, removal of this region has debilitating effects in vivo, including decreased recombination frequency (Steen 1999) and an increase in aberrant recombination (Sekiguchi 2001; Talukder 2004). Interpretation of these effects is complicated by the presence of domains within the RAG-2 non-core region that support destruction of RAG-2 at the G1-S transition (Jiang 2005; Lee and Desiderio 1999; Zhang 2011), govern nuclear import of RAG-2 in cycling cells (Ross 2003) and mediate binding of RAG-2 to H3K4me3 (Liu 2007; Matthews, 2007; Ramon-Maiques 2007). This last interval, which spans residues 415 through 487, comprises a non-canonical plant homeodomain (PHD) finger (Callebaut and Mornon 1998; Ramon-Maiques 2007). In the context of full-length RAG-2, engagement of H3K4me3 by the PHD finger promotes recombination in vivo (Liu 2007; Matthews 2007). Indeed, soluble, synthetic peptides bearing the H3K4me3 modification stimulate cleavage of RSS substrates by RAG in vitro (Grundy 2010; Shimazaki 2009), consistent with the interpretation that H3K4me3 is an allosteric activator of the V(D)J recombinase.

Paradoxically, complete removal of the RAG-2 non-core region impairs V(D)J recombination less severely than does selective mutation of the PHD finger (Cuomo and Oettinger 1994; Sadofsky 1994). Moreover, core RAG-2 supports D-to-J_H joining in vivo (Akamatsu 2003; Kirch 1998; Liang 2002), while the full-length RAG-2(W453A) mutant, which is unable to bind H3K4me3, does so only weakly (Liu et al., 2007). These observations suggest that H3K4me3 imposes allosteric control on V(D)J recombination. By identifying second site mutations that rescue the activity of RAG-2 bearing a

debilitating mutation in the PHD finger we now uncover an autoregulatory region within RAG-2 that imposes allosteric control on V(D)J recombination. We demonstrate that disruption of autoinhibition is associated with constitutive increases in the affinity of RAG for substrate DNA and in the rate of DNA cleavage, thereby mimicking the engagement of H3K4me3. Our observations provide strong support for a model in which epigenetic control of RAG is enforced by an autoinhibitory domain whose action is relieved by active chromatin.

Results

Robust stimulation of RAG activity by exogenous H3K4me3 upon removal of endogenous H3K4me3 (In collaboration with Dr. Chao Lu)

To study the mechanism of interaction between the PHD finger and H3k4me3, we first established a robust system for studying allosteric activation of RAG by H3K4me3. Although exogenous H3K4me3 has been shown to stimulate DNA cleavage by RAG, standard RAG preparations nonetheless support RSS cleavage in the absence of exogenous H3K4me3, and the responsiveness of these preparations to exogenous H3K4me3 is inconsistent. These observations suggested the presence of varying amounts of copurifying, endogenous H3K4me3. Indeed, we discovered that standard amylose affinity preparations of MBP-tagged RAG fusion proteins (Fig 2.1 A) were contaminated with histone H3K4me3. When we employed a protocol that included sonication before amylose affinity chromatography, contaminating H3K4me3 was removed. All RAG preparations employed here were depleted of endogenous H3K4me3 in this way.

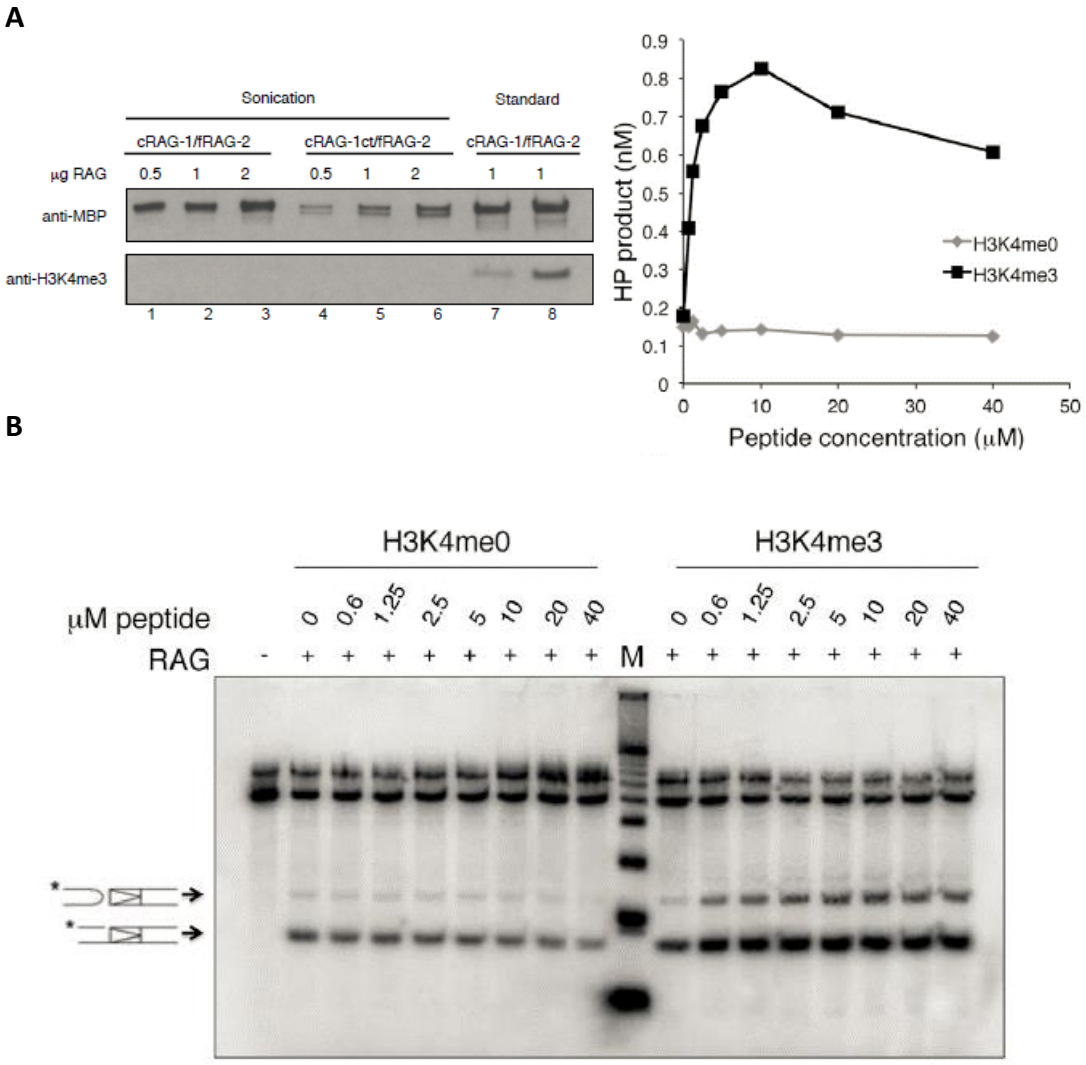


Figure 2.1. A. Separation of RAG from endogenous H3K4me3. Detection of RAG proteins (upper panel) or H3K4me3 (lower panel) by immunoblotting with anti-myc or anti-H3K4me3 antibody, respectively. The RAG fusion proteins cRAG-1-MH (cRAG-1) and fRAG-2-MH (fRAG-2) (lanes 1 – 3, lane 7 and lane 8) or cRAG-1ct-MH (cRAG-1ct) and fRAG-2-MH (lanes 4 – 6). B. Stimulation of H3K4me3-depleted RAG by exogenous H3K4me3. Coupled cleavage reactions contained radiolabeled 12-RSS and unlabeled 23-RSS. Additions of RAG and H3K4me0 or H3K4me3 are indicated above. Positions of hairpin (HP) and nicked (N) products are indicated by arrows. M, 10 bp marker ladder. C. Accumulation of hairpin product (top) or nicked intermediates (bottom) at 1 hr is plotted in nM as a function of H3K4me0 (gray diamonds) or H3K4me3 (black squares) concentration.

We proceeded to assess the effect of exogenous H3K4me3 on DNA cleavage by RAG in the absence of endogenous H3K4me3. In these assays we employed a version of RAG-1 lacking the amino-terminal non-core region but retaining the carboxy-terminal non-core region (cR1ct-MH, Fig 1 C) because RAG complexes containing cR1ct-MH are more robustly stimulated by H3K4me3 than complexes containing canonical core RAG-1 (Grundy 2010), and because we obtained complexes containing cR1ct-MH in at least 20-fold greater yield than complexes containing full-length RAG-1. Full-length RAG-2 (fR2-MH, Fig 1D) and cR1ct-MH fusion proteins were coexpressed and copurified (Fig 2.1 A); the resulting preparation was free of detectable endogenous H3K4me3 (Fig 2.1 A, lanes 4 - 6). This RAG complex was assayed in vitro for coupled cleavage of a radiolabeled 12-RSS in the presence of unlabeled 23-RSS substrate and increasing amounts of a histone H3-derived peptide containing trimethylated lysine 4 (H3K4me3) or unmethylated lysine 4 (H3K4me0). Accumulation of hairpin products was robustly stimulated in a dose-dependent fashion by H3K4me3, but not by H3K4me0 (Fig 2.1 C). Under these conditions the yield of hairpin end-products showed about 4.5-fold maximal stimulation with half-maximal stimulation occurring at 0.6 - 1.25 μM H3K4me3. These results demonstrated the ability of soluble H3K4me3 to stimulate the cleavage of 12-RSS substrates by RAG and suggested that the removal of endogenous H3K4me3 would provide a more well-defined RAG preparation with which to evaluate the effects of exogenous H3K4me3.

Stimulatory effect of H3K4me3 on substrate binding

Our coupled cleavage experiments indicated that H3K4me3 was altering RAG activity by increasing affinity for substrate, increasing the catalytic activity of the enzyme, or both. To distinguish between these possibilities, we independently assessed substrate binding and catalysis with co-purified RAG-1 and RAG-2. To measure affinity for DNA substrate, a canonical 12-RSS fragment was incubated with wild-type RAG at concentrations ranging from 0 to 300 nM in the presence of 4μM H3K4me0 or H3K4me3 peptide. Incubation was carried out in the presence of 5mM Ca^{++}, which supports the binding of RAG to substrate in the absence of DNA cleavage. The fraction of total substrate remaining in the unbound state was determined by gel electrophoresis (Fig. 2.2 A) and expressed as a function of active RAG concentration (Fig 2.2 B), as defined by burst kinetics under the assumption that the active unit is a heterotetramer of composition $(RAG\text{-}1)_2(RAG\text{-}2)_2$ (Yu and Lieber 2000). Dissociation constants, K_D, were determined as described in Experimental Procedures. The addition of H3K4me3 was accompanied by an increase in the affinity of wild-type RAG for substrate DNA, relative to control.

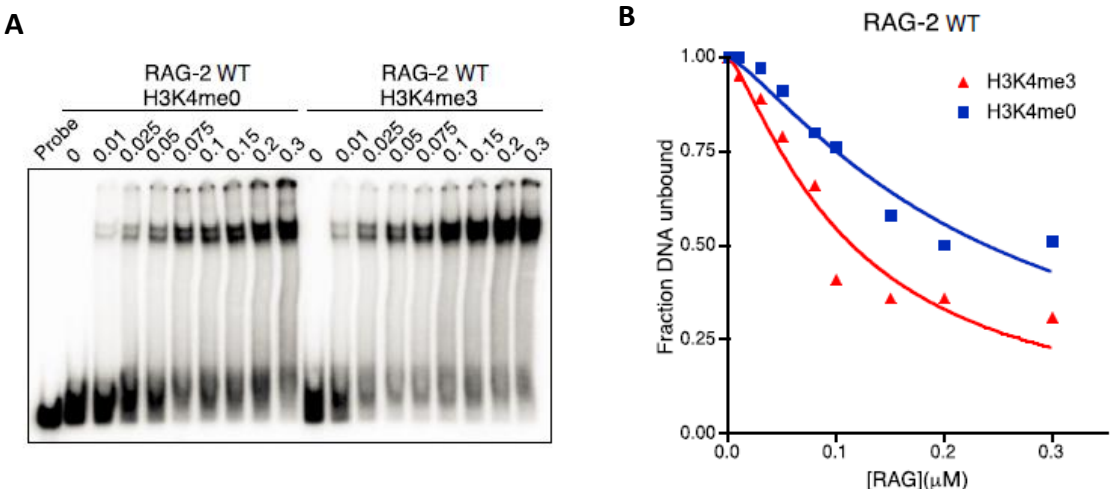


Figure 2.2 Stimulatory effects of H3K4me3 on RAG binding. A. Electrophoretic mobility shift assay (EMSA) for binding of wild-type RAG to a consensus 12-RSS in the presence of H3K4me0 or H3K4me3 peptide as indicated at top. Probe, 12 RSS incubated in the absence of RAG. The concentration, in mM, of active RAG in each binding reaction is indicated above the lane. B. H3K4me3 reduces the KD of RAG-RSS binding. The fraction of free probe (fraction DNA unbound) in each binding reaction of (A) was plotted as a function of active RAG concentration. Data from reactions containing H3K4me0 and H3K4me3 are indicated by blue squares and red triangles, respectively.

In the presence of H3K4me3 wild-type RAG bound substrate with an an estimated K_D of 76 nM, while in the presence of H3K4me0 the K_D was 242nM.

H3K4me3 stimulates catalysis of DNA cleavage

The foregoing experiments suggested that H3K4me3 might have discrete stimulatory effects on substrate binding and catalysis. We therefore subjected RAG-2 in complex with cRAG-1-ct to burst kinetic analysis. Reactions were carried out at an active RAG tetramer concentration of 1.5 nM in the presence of 4 μM H3K4me0 or H3K4me3 peptide (Fig 2.3 A). Following determination of V_{max} (Fig 2.3 A, B), k_{cat} was estimated (Experimental Procedures). We observed turnover rates of 0.83 min^{-1} and 3.76 min^{-1} for wild-type RAG-2 in the presence of H3K4me0 or H3K4me3, respectively (Fig 2.3 B,C), consistent with previous estimates (Shimazaki 2009).

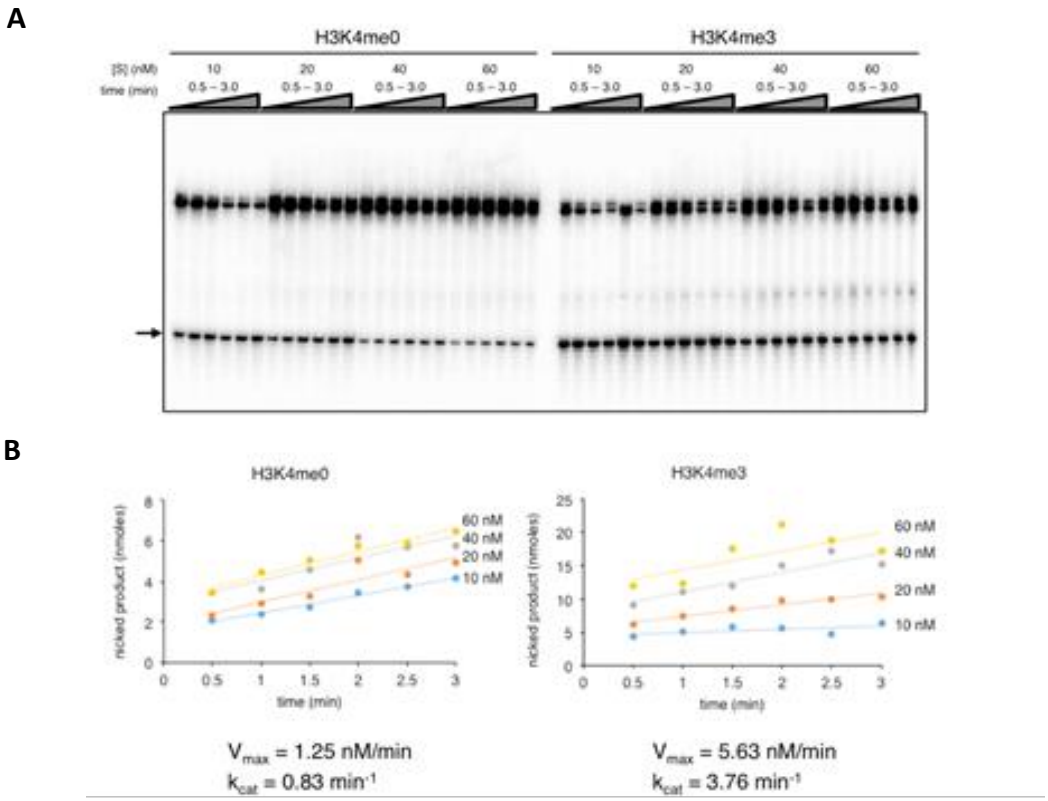


Figure 2.3. Stimulation of RAG DNA nicking by H3k4me3. A. Assay for RSS nicking. Reactions contained 1.5 nM fRAG-2-MH and 12-RSS substrate HL44/45 at 10, 20, 40 or 60 nM. Reactions were supplemented with 4 μM H3K4me0 or H3K4me3 peptide as indicated at top. Accumulation of nicked product (arrow) was assayed at times ranging from 0.5 min to 3 min. B. Concentration of nicked product as determined in (D) is plotted as a function of time for each substrate concentration. Blue, 10 nM; orange, 20 nM; gray, 40 nM; and yellow, 60 nM. Left, reactions containing H3K4me0; right, reactions containing H3K4me3. V_{max} was determined from the slopes of the individual kinetic curves by non-linear regression analysis as described in Experimental Procedures; $k_{cat} = V_{max}/[RAG]_T$, where $[RAG]_T$ is the total concentration of active RAG tetramer.

Taken together these results are consistent with a model in which the RAG activity is suppressed, limiting both substrate binding and catalysis which are overcome upon binding of H3K4me3.

Discussion

The accessibility of antigen receptor loci to V(D)J recombination is tightly associated with transcriptional activity and with epigenetic modifications characteristic of active chromatin. Binding of the RAG-2 PHD finger to one such mark, H3K4me3, is essential for efficient V(D)J recombination in vivo. This has suggested that recombination requires the simultaneous engagement by RAG of appropriately paired RSSs and one or more H3K4me3 moieties. The ability of untethered H3K4me3 to stimulate cleavage of naked DNA by RAG, however, is consistent with the interpretation that H3K4me3 does not simply function as a docking site for the recombinase but rather as an allosteric activator that relieves autoinhibition exerted by some feature of the RAG complex.

Our ability to measure the stimulation of RAG activity by exogenous H3K4me3 relied upon a purification scheme that separated RAG from contaminating, endogenous H3K4me3. The ability of exogenous H3K4me3 to stimulate the coupled cleavage activity of wild-type RAG was dependent, as expected, on binding of H3K4me3 to the PHD finger of RAG-2. H3K4me3 binding exerts at least two effects that contribute to enhanced RSS cleavage activity in vitro: increased affinity of RAG for substrate and a faster catalytic rate. Our results are consistent with a model in which RAG is maintained in a state of low affinity for the RSSs until nearby transcriptional activation promotes allosteric activation through the deposition of H3K4me3. In this view, engagement of the PHD finger by H3K4me3 would relieve inhibition indirectly, perhaps through propagation of a conformational alteration within RAG.

Experimental Methods

Expression constructs

pcDNA3.1 Plasmids encoding full length RAG-2, and plasmids encoding maltose-binding protein (MBP) fusions to full-length RAG-1, core RAG-1 or full-length RAG-2, tagged at the c-termini with a myc epitope and polyhistidine were generated by Alyssa Ward and Chao Lu. To construct a plasmid expressing core RAG-1ct (Grundy 2010), codons 384 through 1040 of mouse RAG-1 were amplified by PCR and ligated in frame between cassettes encoding MBP and myc-polyhistidine that had been inserted into the vector pcDNA3.1.

Burst kinetic analysis

RAG protein was combined at varying nominal concentrations with 200 nM total radiolabeled HL44/45 in binding buffer containing 1% glycerol (reaction volume 10 □l) and incubated for 20 min at 37°C, at which time $MgCl_2$ was added to 5 mM. Incubation was continued for an additional 30 min at 37°C. Reactions were stopped by addition of 10 □l 90% formamide-TBE and heated for 5 min at 95°C. Products were fractioned by electrophoresis on a 15% polyacrylamide-urea gel, visualized by a phosphorimager quantified using ImageQuantNL. The active fraction of each RAG preparation was determined by (1) plotting accumlation of nicked product ([N]) as a function of time for each concentration of total RAG; (2) extrapolating rates to zero time, thereby obtaining the initial burst of nicked product formation ($[N]_0$) at each nominal RAG concentration; and (3) expressing $[N]_0$ as a linear function of nominal RAG concentration:

$$[N]_0 = f_a \cdot [RAG]_T + b$$

where $[RAG]_T$ is the total (nominal) concentration of RAG, assuming a tetrameric stoichiometry of $(RAG\text{-}1)_2(RAG\text{-}2)_2$, and f_a is the fraction of $[RAG]_T$ that is active (Yu and Lieber, 2000).

Assays for coupled cleavage

RAG (1 nM active tetramer, as determined by burst kinetic analysis) was combined in binding buffer with 5 nM HL44/45, 5nM HL46/47 and varying concentrations of a peptide corresponding to the amino-terminal 21 residues of histone H3, either trimethylated at lysine 4 (H3K4me3; Anaspec, 64194) or unmethylated (H3K4me0; Anaspec, 61701), in a reaction volume of 10 µl. After incubation for 20 min at 37°C, $MgCl_2$ was added to 5 mM and incubation was continued for an additional 1 hr. Reactions were stopped and products were analyzed as above.

Assays for DNA nicking by full-length RAG proteins

Full length RAG (1 nM active tetramer, as determined by burst kinetic analysis) was combined in binding buffer with 5 nM HL44/45 and varying concentrations of a peptide corresponding to the amino-terminal 21 residues of histone H3, either trimethylated at lysine 4 (H3K4me3; Anaspec, 64194) or unmethylated (H3K4me0; Anaspec, 61701), in a reaction volume of 10 µl. After incubation for 20 min at 37°C, $MgCl_2$ was added to 5 mM and incubation was continued for an additional 30 min. Reactions were stopped and products were analyzed as above.

Chapter 2

Introduction

All forms of DNA processing – replication, transcription, recombination and repair – employ allosteric regulation, often as a basis for molecular discrimination but also to establish a sequence of interactions or to bias the outcome of a reaction. In many instances the allosteric ligand is a specific DNA structure, as in Cre-mediated recombination, in which the Holliday junction intermediate effects allosteric conformational changes that switch active and inactive Cre monomers (Guo 1997). In other instances the allosteric ligand is a DNA-bound protein array, as in □ integration, which is driven to completion by a flanking DNA-protein array that biases the conformation of λ-integrase (Biswas 2005).

V(D)J recombination, the process by which antigen receptor genes are assembled, is also subject to allosteric control, but in this case the allosteric ligand is a specific chromatin mark rather than a DNA structure. V(D)J recombination is initiated by RAG-1 and RAG-2, which together cleave DNA at recombination signal sequences (RSSs) flanking the participating gene segments (Schatz and Swanson 2011). There are two classes of RSS, termed 12-RSS and 23-RSS, in which heptamer and nonamer elements are separated by spacers of 12 bp or 23 bp; physiological DNA cleavage requires the pairing of a 12-RSS with a 23-RSS (Schatz and Swanson 2011)). V(D)J recombination acts in an ordered, locus-specific fashion during lymphoid development. The accessibility of gene segments to V(D)J recombination is positively correlated with transcription at the unrearranged locus and with histone modifications characteristic of active chromatin, including hypermethylation of histone H3 at lysine 4 (Chakraborty 2007, Jung 2006, Liu 2007, Matthews 2007, Morshead 2003, Subrahmanyam 2012).

RAG-1 and RAG-2 are 1040 and 527 amino acid residues long, respectively. The catalytic core and DNA-binding functions are largely contained within RAG-1 (Kim 2015). RAG-2, which is also essential for DNA cleavage activity, comprises part of a putative DNA-binding cleft and participates in formation of the active site (Kim 2015). Although residues 387 through 527 of RAG-2 are dispensable for DNA cleavage in vitro, this region supports several regulatory functions in vivo, including binding of RAG-2 to H3K4me3 (Liu 2007, Matthews 2007, Ramon-Maiques 2007). This function is mediated by a non-canonical plant homeodomain (PHD) finger that spans residues 415 through 487 (12, 13). Engagement of H3K4me3 by the PHD finger promotes recombination in vivo (Ramon-Maiques 2007, Matthews 2007) and synthetic peptides bearing the H3K4me3 modification stimulate cleavage of RSS substrates by RAG in vitro (Grundy 2010, Lu 2015, Shimazaki 2009), consistent with the interpretation that H3K4me3 is an allosteric activator of the V(D)J recombinase.

This interpretation was reinforced by the identification within RAG-2 of an autoregulatory region, the presence of which was revealed by second site mutations that rescue the activity of RAG-2 lacking a functional PHD finger (Lu 2015). Disruption of this autoregulatory region is associated with constitutive increases in RSS binding affinity, catalytic rate and recombination frequency, thus mimicking the stimulatory effects of H3K4me3 (Lu 2015). These observations support a model in which RAG activity is suppressed by an autoinhibitory domain whose action is relieved by active chromatin.

Allosteric activation is usually accompanied by a change in the distribution of accessible protein conformations in the ligand-bound state (Mortlagh 2014). We asked

whether the conformational distribution of RAG is altered upon binding of H3K4me3 to the RAG-2 PHD finger. To quantify effects of H3K4me3 on the conformations of RAG-1 and RAG-2 we carried out differential isotopic chemical footprinting of solvent-exposed cysteine thiols, in combination with mass spectrometry (Chen 2012). Strikingly, binding of H3K4me3 to RAG-2 is accompanied by robust increases in the solvent accessibility of RAG-1 within the nonamer-binding domain, in a second putative RSS-binding domain and in the ZnH2 domain, which acts as a scaffold for the catalytic center (Kim 2015). These H3K4me3-induced changes in solvent accessibility are abolished by mutation of the RAG-2 PHD finger. Our observations provide direct evidence that engagement of H3K4me3 by the RAG-2 PHD finger is associated with discrete changes in the conformational distribution of RAG-1, consistent with the stimulatory effects of H3K4me3 on RSS binding and cleavage by RAG.

Results and Discussion

Information indicating engagement of H3K4me3 by the RAG-2 PHD finger must be communicated within the RAG tetramer in such a way that substrate affinity and catalytic rate are increased. An allosteric ligand can be thought to act on an ensemble of protein native states by inducing a shift in the energetic distribution of those states and their associated protein conformations (Mortlagh 2014). In accordance with this view, binding of H3K4me3 to the RAG-2 PHD finger may be accompanied by a shift in the distribution of RAG native states, so as to favor conformations associated with increased DNA binding and catalysis. We sought to obtain physical evidence for the induction of such conformational changes by H3K4me3.

Conformational states are expected to differ with respect to the detailed accessibility of amino acid side chains to solvent (Marsh 2011). Solvent-accessible cysteine side chains are susceptible to alkylation by monobromobimane (mBBr), which reacts with free thiols to produce a covalent mBBr-thiol adduct (Chen 2012). We used this reaction, in combination with mass spectrometry, to determine the effect of H3K4me3 binding on the distribution of solvent-exposed cysteine thiols in the RAG tetramer.

The version of RAG-1 employed here, cR1ct, lacks the amino-terminal non-core region and is more soluble than the wild-type protein but retains responsiveness to H3K4me3 in complexes with RAG-2 (Grundy 2010, Lu 2015). In addition to full-length, wild-type RAG-2, we employed RAG-2(W453A), in which the PHD finger cannot bind H3K4me3 (Liu 2007). These RAG-2 variants, tagged at the amino terminus with MBP, were coexpressed individually with MBP-tagged cR1ct and RAG tetramers were purified

by a protocol that removes endogenous H3K4me3 (15) (Fig 3.1A). The active fractions of these RAG preparations, as determined by burst kinetics (Fig 3.1B – C) ranged from

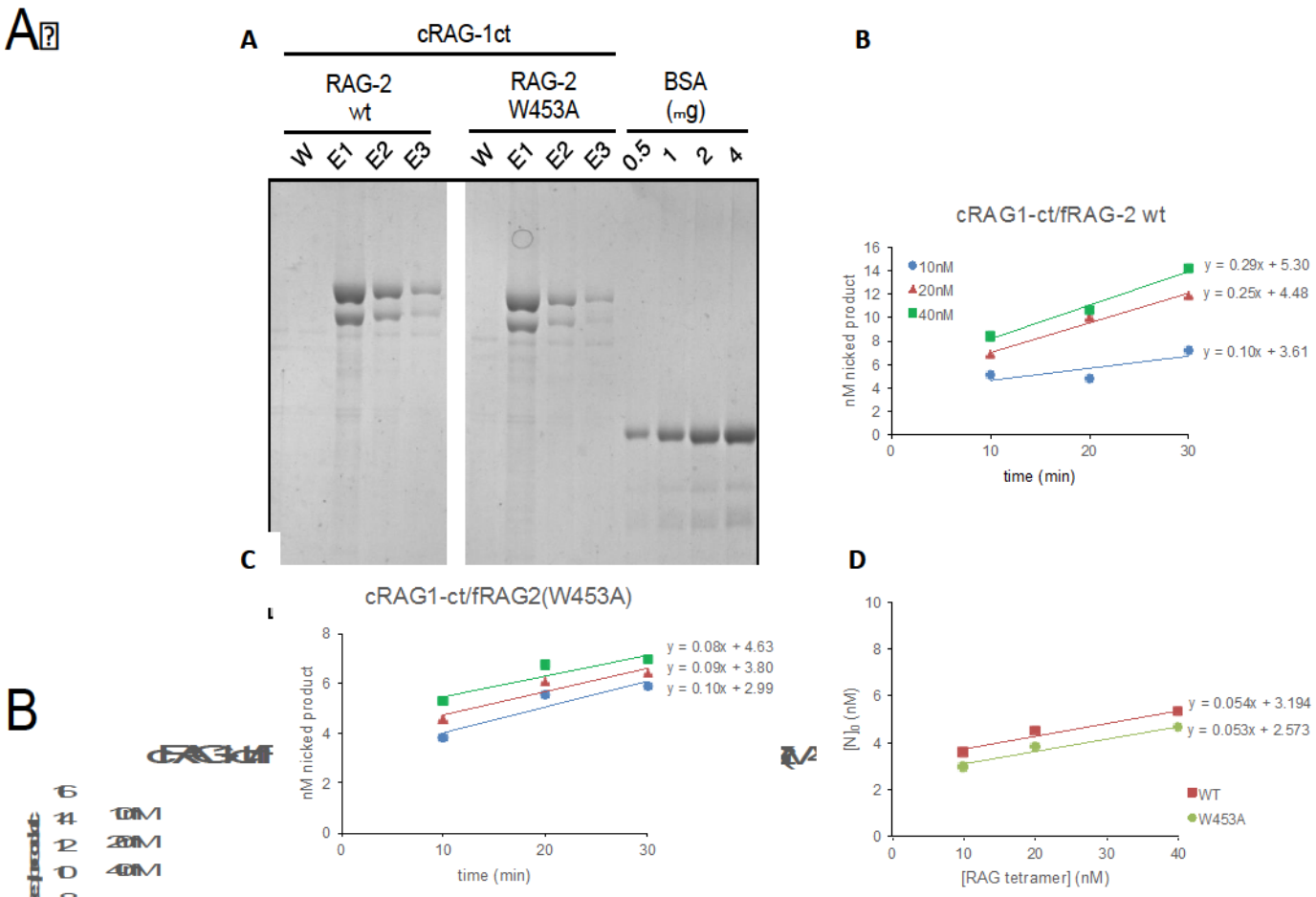


Figure 3.1 A. Purification of wild-type and mutant RAG proteins. (A) Complexes of cRAG-1-ct with full-length wild-type RAG-2 (WT) or RAG-2(W453A) were purified by sonication and amylose affinity chromatography. Aliquots of each fraction were separated by SDS/PAGE alongside a dilution series of BSA.W denotes wash; E1, E2, and E3 denote sequential amylose eluates. All samples were run on the same gel. B. Burst kinetic analysis of wild-type RAG. A complex of cRAG-1ct with full-length, wild-type RAG-2 was assayed for nicking of a 12-RSS substrate at various nominal RAG concentrations. Accumulation of nicked product is plotted as a function of time at each nominal RAG concentration as indicated in the inset. C. Assay of 12-RSS nicking by full-length RAG-2(W453A) in complex with cRAG- 1-ct, performed at various nominal RAG concentrations and displayed as in B. D. Estimation of the active fraction. Slope of the line is fraction of active protein.

Equivalent amounts of each active RAG tetramer were assayed in vitro for couple cleavage of a radiolabeled 12-RSS in the presence of unlabeled 23-RSS substrate and increasing amounts of a histone H3-derived peptide containing trimethylated lysine 4 (H3K4me3) or unmethylated lysine 4 (H3K4me0).

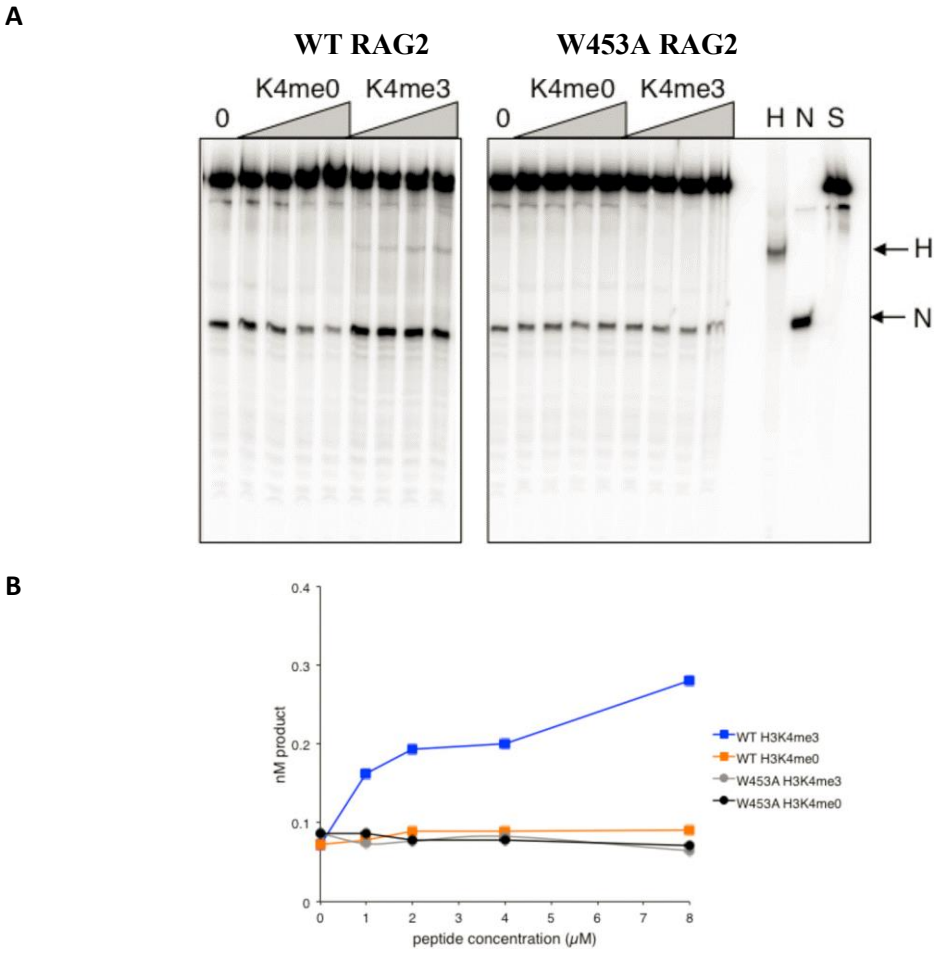


Figure 3.2 H3K4me3 stimulates coupled DNA cleavage by RAG complexes containing wild-type RAG-2 but not RAG-2(W453A). A. Coupled cleavage reactions contained radiolabeled 12-RSS and unlabeled 23-RSS. Reactions contained wild-type (WT) RAG-2 or RAG-2(W453A) as defined at top. K4me0 and K4me3, reactions supplemented with 0.5, 1, 2 or 4 μM H3K4me0 or H3K4me3 peptide; 0, reactions lacking peptide. Markers corresponding to intact substrate (S) hairpin product (H), or nicked product (N) are indicated at right. All samples were run on the same gel. B. Accumulation of hairpin product at 1 h (nanomolar product) is plotted as a function of the concentration of H3K4me0 or H3K4me3. Blue quares, wild-type RAG-2 with H3K4me3; orange squares, wild-type RAG-2 with H3K4me0; gray circles, RAG-2(W453A) with H3K4me3; and black circles, RAG-2(W453A) with H3K4me0.

As previously reported (Lu 2015), the activity of wild-type RAG was stimulated in a dose-dependent fashion by H3K4me3, but not by H3K4me0, while the RAG-2(W453A) mutation abolished responsiveness to H3K4me3 (Fig. 3.2A, B). We proceeded to employ these preparations in chemical footprinting and partial proteolysis assays.

Mapping solvent accessibility of RAG cysteine thiols by differential isotopic chemical footprinting

The chemical footprinting procedure is outlined in Fig 3.3 A. Two protein conformations are depicted, one in the absence (Fig 3.3 A, top left) and one in the presence (Fig 3.3 A, top right) of bound H3K4me3; these single conformations are meant to represent populations of conformations whose distributions differ. RAG preparations were pre-incubated in the presence of H3K4me3 or H3K4me0 and then subjected to a 1 minute pulse of alkylation with light mBBr ([^{1}H]mBBr). After alkylation by [^{1}H]mBBr was quenched, RAG was denatured in guanidinium chloride (GdmCl) and unmodified cysteine thiols were alkylated with hexadeuterated mBBr-d6 ([^{2}H]mBBr). Light alkylation is represented in Fig. 3.3 A by red plain R; heavy alkylation by red boldface R. The modified RAG subunits were digested with trypsin and the products were analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS). Idealized mass spectrograms are diagrammed at the bottom of Fig 3.3 A, where the x axes denote m/z, the y axes denote chromatographic retention time and the z axes denote relative peak intensity. The mass spectra on the left depict the theoretical distribution of ^{13}C isotopologues for a tryptic peptide in which a specific cysteine residue, C_P, has been

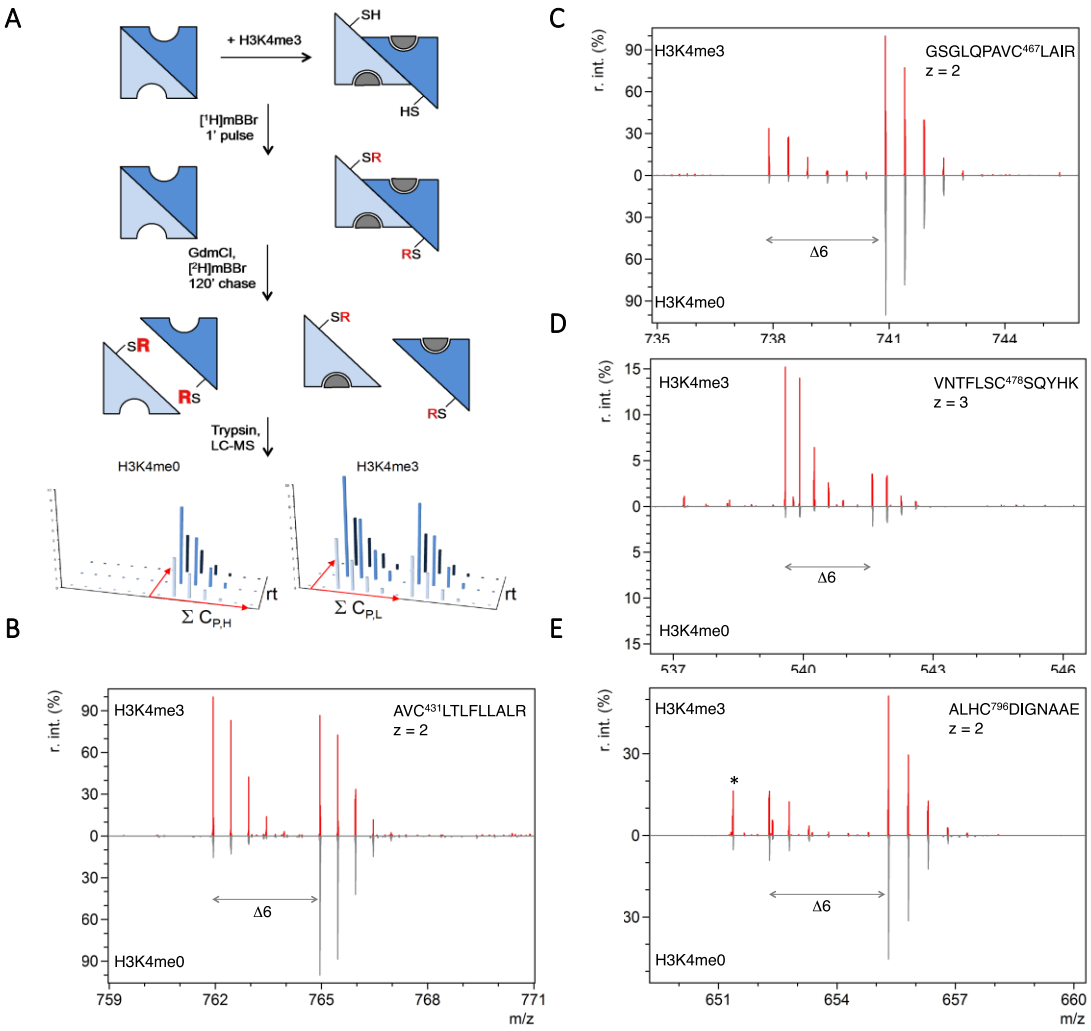


Figure 3.3. Mapping H3K4me3-induced conformational changes in RAG by pulse alkylation mass spectrometry. A. Experimental scheme. Binding of H3K4me3 is proposed to result in a shift in the distribution of RAG conformations, accompanied by a change in the accessibility of some cysteine thiols to solvent. Solvent accessibility is probed by pulse alkylation with [1H]mBBr. Following quenching of the first alkylation reaction and denaturation of RAG in guanidinium chloride (GdmCl), remaining unmodified cysteine thiols are alkylated with [2H]mBBr. RAG is then fragmented with trypsin and peptides resolved by liquid chromatography-mass spectrometry (LC-MS). For each cysteine-containing peptide, total light or heavy alkylation (ΣCP,L and ΣCP,H, respectively) is determined by summing the corresponding peak areas over two dimensions (red arrows): isotopologues and retention time (rt). B – E. Representative spectra of four mBBr-modified peptides from RAG-1. Percent relative intensity is plotted as a function of mass:charge ratio. Spectra obtained from H3K4me3-treated samples are shown above the midline in red; spectra obtained from H3K4me30-treated samples are reflected below the midline in gray. Sequences of the corresponding peptides and their charges (z) are given at upper right. The heavy-modified peak is separated from the light-modified peak by 6 daltons (Δ6). Images of mass spectra were generated in mMass Niedermeyer et al., 2012).

alkylated by mBBr-d6. The mass spectra on the right indicate the ^{13}C isotopologues of the same tryptic peptide, in which C_P has been alkylated by light mBBr or mBBr-d6. Such a pattern would be expected if the addition of H3K4me3 were to alter the conformational distribution of RAG so as to increase the accessibility of residue C_P to solvent.

Tryptic peptides containing cysteine residues alkylated by light mBBr were identified by a mass shift of 190.20 Da relative to the unmodified form; peptides modified by mBBr-d6 exhibited the expected mass shift of 6.05 relative to their light modified counterparts. Differences in the patterns of mBBr alkylation obtained in the presence of H3K4me3 or H3K4me0 were observed for several cysteine-containing peptides from RAG-1 (Fig 3.3 B – E). For example, a tryptic peptide spanning residue C431 exhibited robust alkylation by light mBBr in the presence of H3K4me3 (Fig 3.3 B, upper panel, red) but was nearly refractory to alkylation by light mBBr in the presence of H3K4me0 (Fig 3.3 B, lower panel, gray). The extent of alkylation by heavy mBBr following denaturation of RAG was similar in the presence of H3K4me0 or H3K4me3 (Fig 3.3 B). Representative mass spectra corresponding to peptides spanning C467, C478 and C796 of RAG-1 also exhibited apparent differences between the H3K4me3 and H3K4me0 spectra with respect to alkylation by light mBBr (Fig 3.3 C – E).

Differential footprinting with mBBr and mBBr-d6 provided a robust means to quantify the extent of pulse alkylation and to compare the efficiencies of pulse alkylation in samples that had been treated with H3K4me3 or H3K4me0. This was achieved in three steps. First, spectra representing all detectable peptides spanning a particular cysteine residue C_P were binned. Second, the peak intensities for all tryptic peptides

containing C_P were summed over retention time and over the ^{13}C isotopologues corresponding to heavy or light alkylation ($\Sigma C_{P.H}$ and $\Sigma C_{P.L}$, respectively; Fig 3.3 A, bottom). Third, the relative accessibility of each modified cysteine residue to light mBBr was defined as $L_C/(L_C + H_C)$, where L is $\Sigma C_{P.L}$ and H is $\Sigma C_{P.H}$. Three independent pulse alkylation-mass spectrometry assays were carried out for wild-type RAG and the mutant RAG preparation.

The relative accessibility of RAG-1 cysteine residues to pulse alkylation in a native protein preparation incubated with the H3K4me0 control showed good correlation with the computed solvent-accessible area of cysteine residues in two available RAG structures (Kim 2015,Ru 2015) (Fig 3.4). We noted that the accessibility of RAG-1 C431 to pulse alkylation was more closely correlated with its solvent-accessible area in the electron microscopic structure (Ru 2015) than in the crystallographic structure (Kim 2015), suggesting that the former structure may more closely approximate the predominant solution conformation of the NBD in the absence of H3K4me3

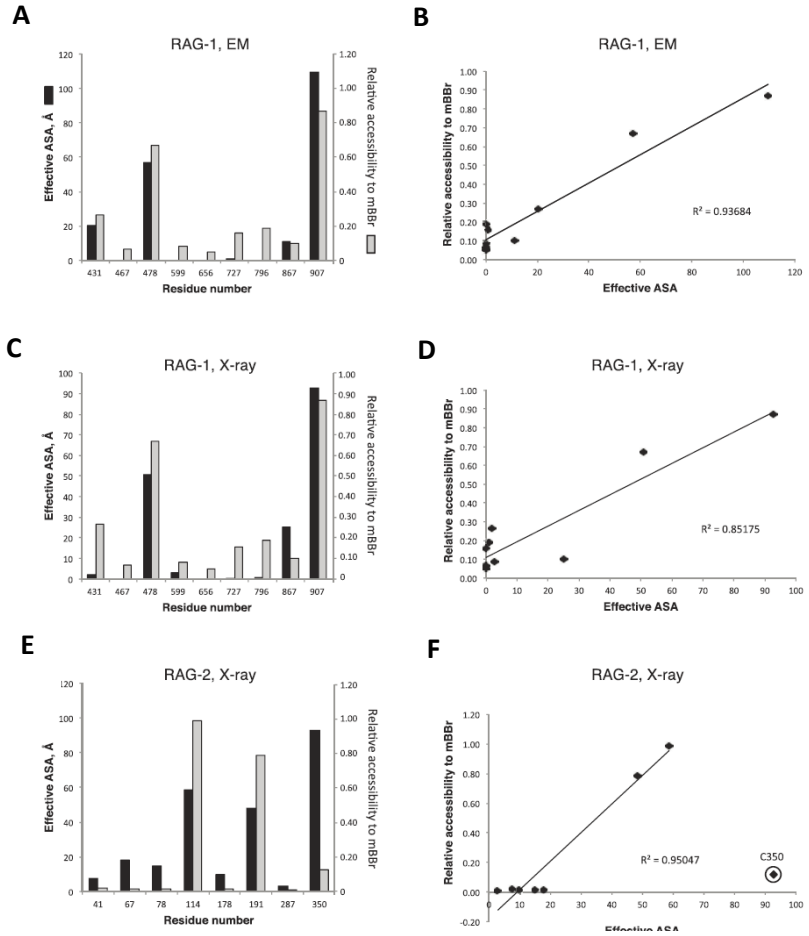


Figure 3.4. Accessibility of RAG cysteine thiols to mBBr compared with accessible surface areas computed from structural data. A. Effective accessible surface areas (ASA) of cysteine residues in RAG-1 (dark gray bars) were computed from a cryoelectron microscopic structure of RAG (PDB ID code 3JBW) (20) using PISA as described in Experimental Methods. Relative accessibility to mBBr (light gray bars) was determined experimentally and is equivalent to L/(L + H) as defined in Experimental Methods. B. Correlation between effective ASA as computed in A with relative accessibility to mBBr. C. Effective accessible surface areas (ASA) of cysteine residues in RAG-1 (dark gray bars) were computed from an X-ray crystallographic structure of RAG (PDB ID code 4WWX) (11) using PISA as described in Experimental Methods. Relative accessibility to mBBr (light gray bars) is as defined in A. D. Correlation between effective ASA as computed in C with relative accessibility to mBBr. E. Effective ASA of cysteine residues in RAG-2 (dark gray bars) were computed from an X-ray crystallographic structure of RAG (PDB ID code 4WWX) (11) using PISA as described in Experimental Methods. Relative accessibility to mBBr (light gray bars) was determined experimentally and is equivalent to L/(L + H) as defined in Experimental Methods. F. Correlation between effective ASA as computed in E with relative accessibility to mBBr. The point corresponding to residue C350 (circled) was excluded from analysis because its high effective ASA may be an artifact of the truncation of RAG-2 that was used in determining the X-ray structure.

(Fig 3.4, A – D). Alkylation of RAG-2 cysteine residues could only be compared to the solvent accessible areas computed from the crystallographic structure, because the electron microscopic structure employed zebrafish RAG-2, which is highly divergent from murine RAG-2. The only large discrepancy we observed between accessibility to pulse alkylation and computed effective surface area was at C350 of RAG-2, which was inaccessible to mBBr in the native state but exposed in the crystal structure (Fig 3.4, E and F). This difference is likely explained by the fact that the X-ray structure was determined using a truncated form of RAG-2 in which the carboxy-terminal residues beyond C350 were disordered, while our pulse alkylation assays were performed with full-length RAG-2.

H3K4me3 binding to the RAG-2 PHD finger is associated with solvent accessibility changes in RAG-1

For RAG-1 in complex with wild-type RAG-2, peptides spanning 10 of the 15 cysteine residues in cR1ct were consistently detectable. For each of these peptides, the relative accessibility, $L_C/(L_C + H_C)$, of each detectable cysteine residue in the presence of H3K4me3 (Fig 3.5 A, red bars) was compared to its relative accessibility in the presence of H3K4me0 (Fig 3.5 A, gray bars). Significant increases in accessibility of cysteine to alkylation by mBBr were observed at four positions: 431, 467, 478 and 796. No significant decreases in relative accessibility were associated with H3K4me3 treatment. The effect of H3K4me3 on accessibility is depicted alternatively in Fig 3.5 B as the difference between the mean relative accessibility measured in the presence of

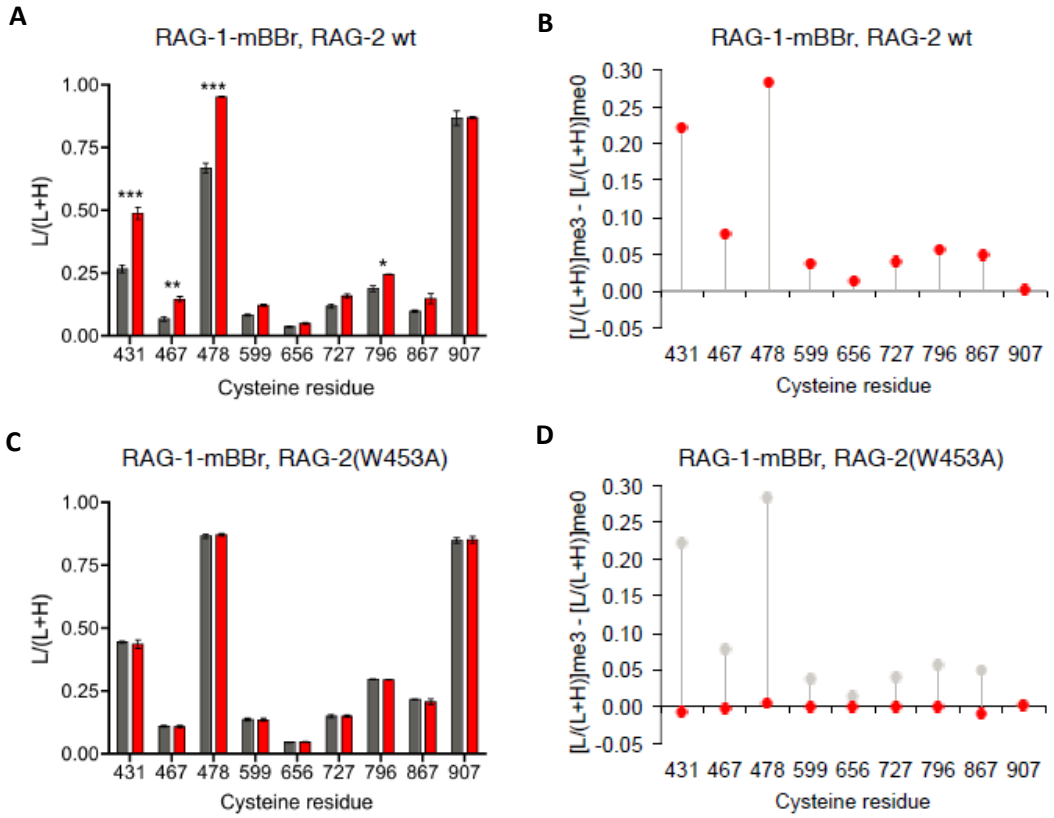


Figure 3.5. Binding of H3K4me3 to the PHD of RAG-2 induces localized conformational changes in RAG-1. A. Relative accessibility [L/(L+H)] of RAG-1 cysteine thiols to pulse alkylation by light mBBr in the presence of wild-type RAG-2 and H3K4me3 (red) or H3K4me0 (gray). Residues are numbered below. Values indicate mean and S.E.M. of three trials. Statistical significance was determined by ANOVA. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. B. Differential mean relative accessibilities ([L/(L+H)]me3 - [L/(L+H)]me0) of RAG-1 cysteine thiols in wild-type RAG complexes pulse alkylated in the presence of H3K4me3 or H3K4me0. Residues are numbered below. C. Relative accessibility [L/(L+H)] of RAG-1 cysteine thiols to pulse alkylation by light mBBr in the presence of RAG-2(W453A) and H3K4me3 (red) or H3K4me0 (gray). Residues are numbered below. Values indicate mean and S.E.M. of three trials. Statistical significance was determined by ANOVA and indicated as in (*A*). D. Differential mean relative accessibilities ([L/(L+H)]me3 - [L/(L+H)]me0) of RAG-1 cysteine thiols in complexes containing RAG-2(W453A), pulse alkylated in the presence of H3K4me3 or H3K4me0.

H3K4me3 ([L_C/(L_C + H_C)]me3) and that measured in the presence of H3K4me0 ([L_C/(L_C + H_C)]me0). These results indicate that several cysteine thiols in RAG-1 exhibit increased susceptibility to alkylation by mBBr in the presence of H3K4me3. H3K4me3 exerts its allosteric effects on DNA binding and catalysis through an interaction with the PHD finger of RAG-2. To determine whether the effects of H3K4me3 on alkylation of RAG-1 are similarly mediated, we performed differential footprinting with RAG complexes containing the RAG-2 W453A mutation, which disrupts binding of the PHD finger to H3K4me3. The W453A mutation abolished the ability of H3K4me3 to enhance alkylation by mBBr at each of the positions where enhanced alkylation was observed in the wild-type complex (Fig 3.5 C and D). Taken together these observations indicate that engagement of H3K4me3 by the PHD finger of RAG-2 is communicated to the RAG-1 subunit as an alteration in the distribution of RAG-1 conformations.

H3K4me3-induced changes in mBBr accessibility not detected in the RAG-2 core region

Peptides spanning 11 of the 17 cysteine residues in full-length, wild-type RAG-2 were consistently detected. Of these cysteine residues, eight lie within the RAG-2 core region and three within the PHD finger; residues from the highly acidic autoinhibitory region were inconsistently detected. Residues C114, C191 and to a lesser extent C350 were pulse alkylated by light mBBr in the absence of H3K4me3; no significant changes in the extent of alkylation at these or other cysteine residues were observed in the presence of H3K4me3 (Fig 3.6A, B). The W453A mutation had no effect on the accessibility of RAG-2 core residues to mBBr in the presence of either H3K4me3 or H3K4me0, but was associated with a large increase in basal alkylation of C446 and

C478, which reside within the PHD finger. The apparent increase in solvent accessibility of these residues is consistent with a destabilizing effect of the W453A mutation on the PHD finger (Fig 3.6 C, D).

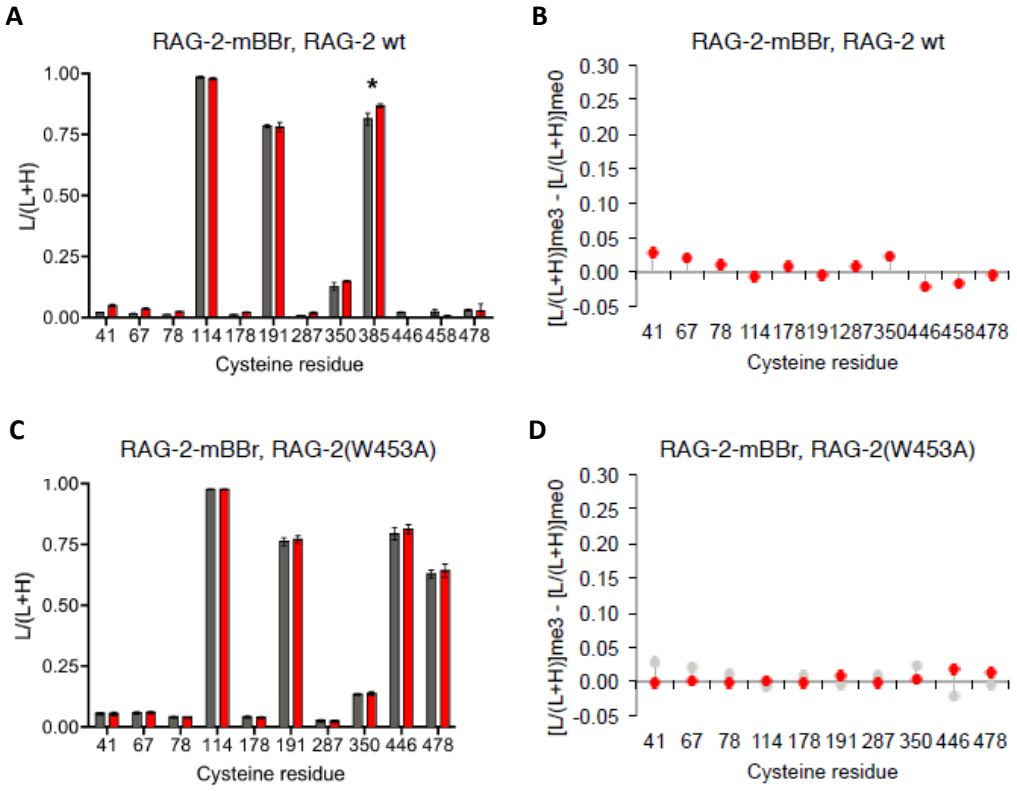


Figure 3.6. Accessibility of RAG-2 cysteine thiols to mBBr in the presence of H3K4me3 or H3K4me0. (A) Relative accessibility [L/(L+H)] of wild-type RAG-2 cysteine thiols to pulse alkylation of RAG tetramers by light mBBr in the presence of H3K4me3 (red) or H3K4me0 (gray). Residues are numbered below. Values indicate mean and S.E.M. of three trials. No significant differences (ANOVA) were observed at the cysteine residues detectable by LC-MS. (B) Differential mean relative accessibilities ([L/(L+H)]me3 - [L/(L+H)]me0) of RAG-2 cysteine thiols in wild-type RAG complexes pulse alkylated in the presence of H3K4me3 or H3K4me0. Residues are numbered below. (C) Relative accessibility [L/(L+H)] of RAG-2(W453A) cysteine thiols to pulse alkylation of RAG tetramers by light mBBr in the presence of H3K4me3 (red) or H3K4me0 (gray). Residues are numbered below. Values indicate mean and S.E.M. of three trials. No significant differences (ANOVA) were observed at the cysteine residues detectable by LC-MS. (D) Differential mean relative accessibilities ([L/(L+H)]me3 - [L/(L+H)]me0) of RAG- 2(W453A) cysteine thiols in RAG complexes pulse alkylated in the presence of H3K4me3 or H3K4me0.

We considered the possibility that the alterations in cysteine alkylation induced by H3K4me3 might result from changes in the association of RAG-1 with RAG-2, rather than alterations in the conformation of tetrameric RAG. One argument against a H3K4me3-dependent change in the distribution of RAG multimers, however, comes from a consideration of the pulse alkylation data for RAG-2. The available RAG structures (11, 20) locate Cys41 and Cys67 of RAG-2 at an interface with RAG-1. We detected no H3K4me3-dependent solvent accessibility changes at either of these residues, as would have been expected from an alteration in the association of RAG-2 with RAG-1, and both residues remain relatively inaccessible to mBBr in the presence of H3K4me0 or H3K4me3 (Fig 3.6 A, B). Two additional lines of evidence support this interpretation. First, the active fraction of wild-type RAG, as assessed by burst kinetic analysis, was similar in the presence of H3K4me0 or H3K4me3 (5.23% and 5.25%, respectively; Fig 3.7).

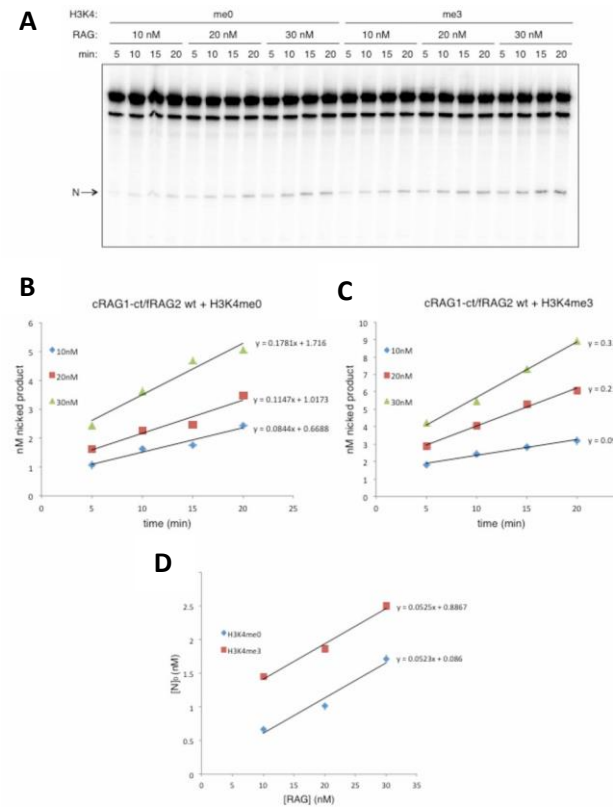


Figure 3.7 H3K4me3 does not alter the active fraction of RAG as determined by burst kinetic analysis. A. Burst kinetic analysis of wild-type RAG. A complex of cRAG-1ct with full-length, wild-type RAG-2 was assayed for nicking of a 12-RSS substrate at various nominal RAG concentrations in the presence of H3K4me3 or H3K4me0. Accumulation of nicked product was assayed at 5, 10, 15, and 20 min. N, nicked product. B. Accumulation of nicked product in the presence of H3K4me0 is plotted as a function of time at each nominal RAG concentration as indicated in the inset. C. Accumulation of nicked product in the presence of H3K4me3 is plotted as a function of time at each nominal RAG concentration as indicated in the inset. D. Estimation of the active fraction. Kinetic curves of B and C were extrapolated to zero time to give the [N]0 associated with each nominal RAG concentration ([RAG]). [N]0 was then plotted as a function of [RAG]; the slope of each resulting curve represents the fraction of active RAG in the preparation.

Second, electrophoretic mobility shift assays typically resolve two discrete RAG-DNA complexes, each containing a RAG-1 dimer and one or two RAG-2 subunits (Swanson 2002); we observed no difference in the relative yields of these complexes in the presence of H3K4me0 or H3K4me3 (Fig 3.8). These observations suggest that under the conditions of these assays the introduction of H3K4me3 is unaccompanied by an alteration in the amount of RAG-1 associated with RAG-2.

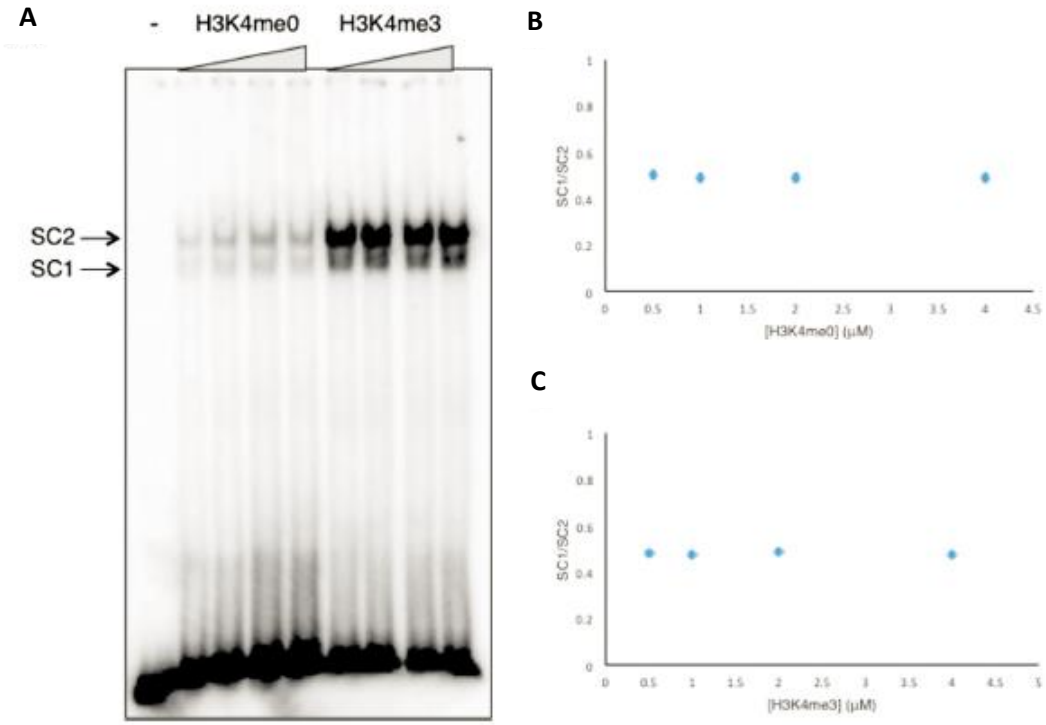


Figure 3.8 H3K4me3 does not alter the distribution of RAG–RSS complexes as assessed by electrophoretic mobility shift. A. Electrophoresis mobility shift assay (EMSA) for wild-type RAG-2/cRAG1-ct binding to a radiolabeled consensus 12-RSS DNA substrate, in the presence of increasing concentrations of H3K4me0 or H3K4me3 peptide. RAG (0.3 μM) was incubated with 1 nM 12-RSS probe and H3K4 peptide (0.5–4 μM) for 30 min at 37 °C. After addition of 4 μL 50% (vol/vol) glycerol, reactions were resolved on a 4% (vol/vol) acrylamide gel. B. Ratio of SC1:SC2 as a function of H3K4me0 concentration. C. Ratio of SC1:SC2 as a function of H3K4me3 concentration.

The active fraction of the RAG preparations employed above, as assessed by burst kinetic analysis, was only about five percent, likely a result of the harsh sonication employed to remove H3K4me3. We sought a way to obtain a more active preparation of RAG with which to assess the effects of H3K4me3 on conformation. By avoiding sonication and introducing a mild DNase treatment prior to amylose affinity purification, we were able to obtain RAG free from detectable H3K4me3 and residual DNase activity

(Fig 3.9, A and B). The activity of this preparation, as assessed by burst kinetic analysis, was about 24 percent (Fig 3.9, C and D).

This more active fraction exhibited H3K4me3-induced changes in mBBr accessibility similar to those observed for the less active fraction (Fig 3.9 E), indicating that the conformational alterations we detected with the sonicated preparations were not an artifact of the mode of purification. This result, moreover, suggests that some portion of inactive RAG, as defined by burst kinetic analysis, retains the ability to undergo conformational change in response to H3K4me3.

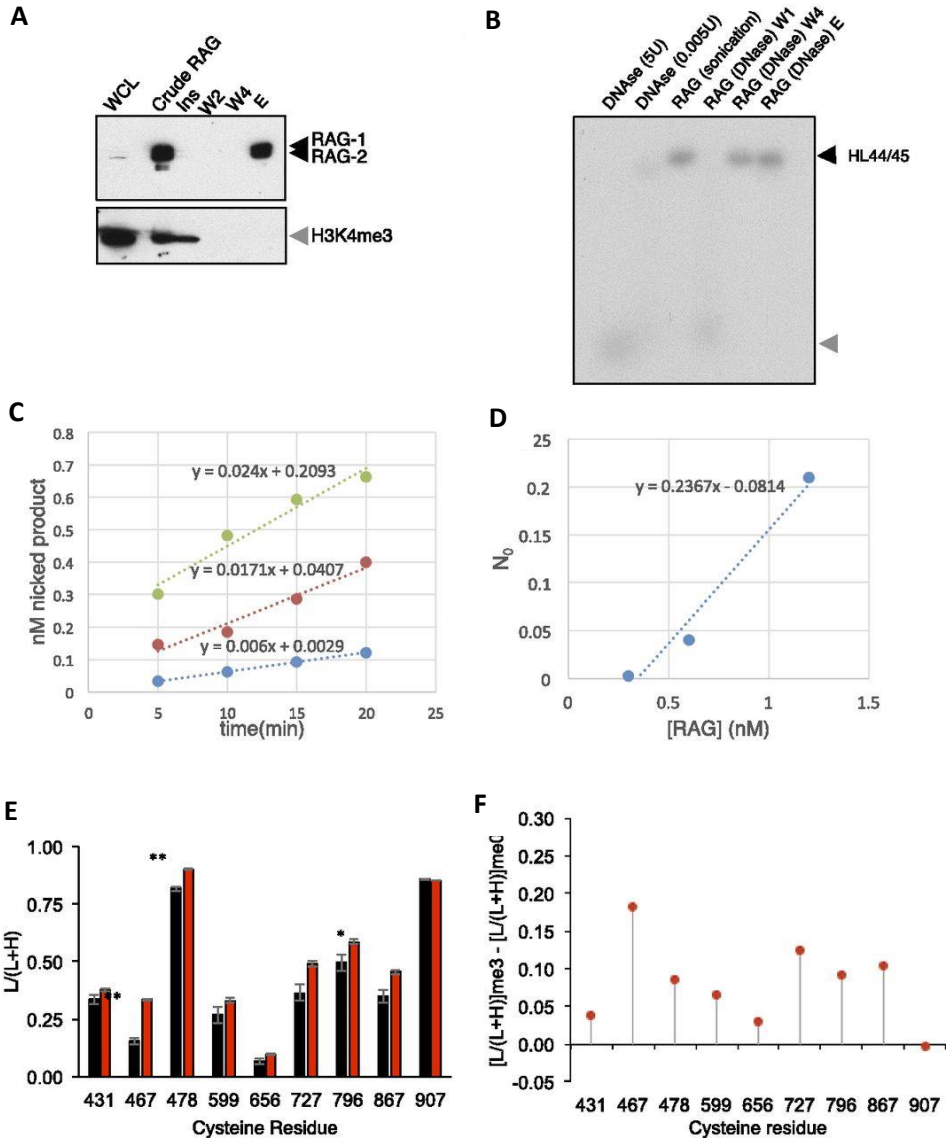


Figure 3.9 Assay for altered solvent accessibility in RAG prepared by a gentle method. (A) Assay for RAG-1, RAG-2, and H3K4me3 at various stages of purification by immunoblotting. (Upper) Detection of RAG using an anti-MBP antibody. (Lower) Immunoblotting with an anti-H3K4me2/3 antibody. E, eluate; Ins, insoluble fraction; WCL, whole-cell lysate; W2, wash 2; and W4, wash 4. (B) Purified RAG is free of detectable DNase contamination. Selected RAG purification fractions were incubated with radiolabeled substrate HL44/45 for 1 h at 37 °C. DNase was assayed in the indicated amounts and a standard sonicated RAG preparation was used as a negative control. E, eluate; W1, wash 1; and W4, wash 4. (C and D). Burst kinetic analysis of gently purified RAG. Active fraction was determined to be 23.6%. (E) Relative accessibility [L/(L + H)] of RAG-1 cysteine thiols to pulse alkylation by light mBBr in the presence of gently purified wild-type RAG-2 and H3K4me3 (red) or H3K4me0 (gray). Residues are numbered below. Values indicate mean and SEM of three trials. Statistical significance was determined by ANOVA. *$P < 0.05$; **$P < 0.01$. (F) Differential mean relative accessibilities ([L/(L + H)]me3 − [L/(L + H)]me0) of RAG-1 cysteine thiols, pulse alkylated in the presence of H3K4me3 or H3K4me0.

H3K4me3-induced conformational changes in RAG-1 and RAG-2 mapped by limited proteolysis and LC-MS/MS

We used limited proteolysis in combination with LC-MS/MS to confirm that H3K4me3 binding to the RAG-2 PHD finger is associated with conformational changes in RAG. We chose thermolysin for this purpose because of its relatively low specificity, which would increase the likelihood of identifying cleavage sites whose accessibility was altered upon binding of H3K4me3. Moreover, the use of thermolysin was expected to promote detection of peptides from regions with a paucity of tryptic cleavage sites, in particular the RAG-2 autoinhibitory domain. RAG complexes containing wild-type RAG-2 or RAG-2(W453A) were incubated with thermolysin for various times in the presence of H3K4me3 or H3K4me0. The mass distribution of partial proteolytic fragments obtained from wild-type RAG in the presence of H3K4me3 differed from the distribution obtained in the presence of H3K4me0, in that H3K4me3 enhanced production of fragments migrating at about 85 kD and 60 kD (Fig 3.10, arrows). In contrast, the proteolytic pattern observed for the mutant RAG complex in the presence of H3K4me3 was similar to the pattern obtained in the presence of H3K4me0 (Fig 3.10 A), indicating that the effect of H3K4me3 on proteolysis was dependent on an interaction with the PHD finger. The fragments whose production was stimulated by H3K4me3 were excised, subjected to digestion with Lys-C and identified by LC-MS/MS. Sites of thermolysin cleavage were determined by comparison to reference spectra obtained for RAG digested with Lys-C alone. The 85 kD band was produced by thermolysin cleavage of RAG-2 at two neighboring sites, F384 and F386 (Fig 3.10 A, B). The resulting amino-terminal fragments, including the MBP moiety, have calculated molecular weights of

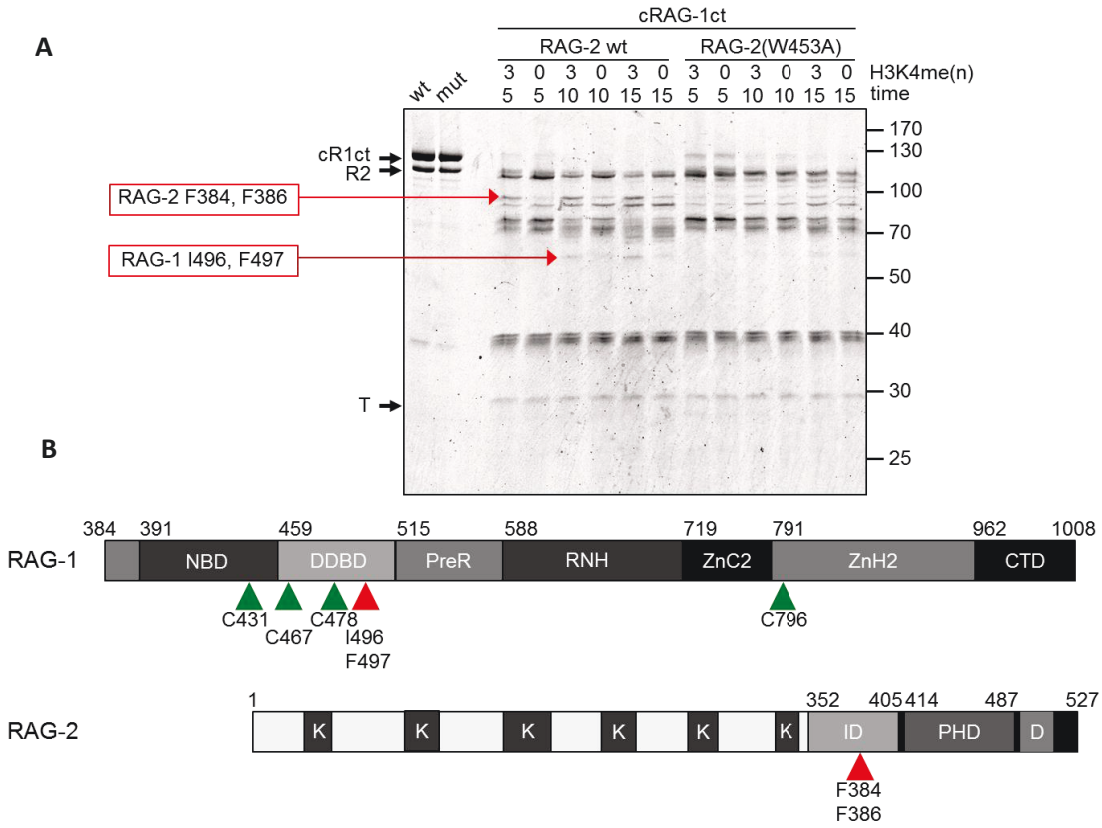


Figure 3.10 H3K4me3 binding to the RAG-2 PHD exposes thermolysin cleavage sites in a DNA-binding domain of RAG-1 and the autoinhibitory domain of RAG-2. (A) RAG tetramers containing cRAG-1ct in complex with wild-type RAG-2 or RAG-2(W453A) were incubated with thermolysin in the presence of H3K4me3 or H3K4me0. H3K4me(n), H3K4 methylation state (3 or 0). Time of incubation is indicated in minutes. cR1ct and R2 indicate the positions of intact cRAG-1ct and RAG-2, respectively. Red arrows indicate the positions of cleavage products whose excision from wild-type RAG-2-containing complexes is stimulated by H3K4me3; sites of thermolysin cleavage as determined by LC-MS are indicated. (B) Sites of H3K4me3-stimulated alkylation (green triangles) and proteolysis (red triangles) are superimposed on diagrams of RAG-1 (upper panel) and RAG-2 (lower panel). NBD, nonamer-binding domain; DDBD, dimerization and DNA-binding domain; PreR, pre-RNaseH domain; RNH, RNaseH domain; ZnC2, ZnC2 domain; ZnH2, ZnH2 domain; CTD, carboxy-terminal domain; K, Kelch-like domain; ID, autoinhibitory domain; PHD, plant homeodomain; D, degradation domain. Residues at domain boundaries are numbered above.

84.49 and 84.74 kD, in good agreement with their electrophoretic mobilities. The 60 kD species was produced by thermolysin cleavage of RAG-1 at I496 and F497 (Fig 3.10 A,

B), which is predicted to yield carboxy-terminal fragments of 62.21 and 62.32 kD, once again consistent with the observed mobilities.

Localization of H3K4me3-induced conformational changes to DNA-binding, catalytic and autoinhibitory domains of RAG

RAG-1 contains six discrete structural domains (Kim 2015,Yin 2009). Of the four cysteine residues of RAG-1 that exhibit increased solvent accessibility upon H3K4me3 binding, three reside in domains responsible for RSS recognition: C431 lies in the nonamer binding domain (NBD) (Fig. 5A and C) while C467 and C478 lie in the dimerization and DNA-binding domain (DDBD) (Fig. 5B and D). The fourth residue, C796, resides in the ZnH2 domain (Fig. 5A and E), which acts as a scaffold for the catalytic region (Kim 2015). The sites in RAG-1 at which H3K4me3 enhances thermolysin cleavage, I496 and F497, also reside in the DDBD, providing additional evidence for an allosteric conformational change in this domain (Fig 3.10 B). F384 and F386 lie within the acidic autoinhibitory domain of RAG-2 (3.10 B), which is not contained within the available structures (Kim 2015, Ru 2015, Yin 2009). The ability of H3K4me3 to promote proteolysis at these sites indicates that allosteric activation of RAG-2 is accompanied by increased accessibility of the autoinhibitory domain to thermolysin. This change in conformation may reflect disruption of an inhibitory interaction that suppresses RAG activity, as we previously demonstrated that mutation of the RAG-2 autoinhibitory domain mimics the stimulatory effects of H3K4me3 on RSS binding and catalysis of DNA

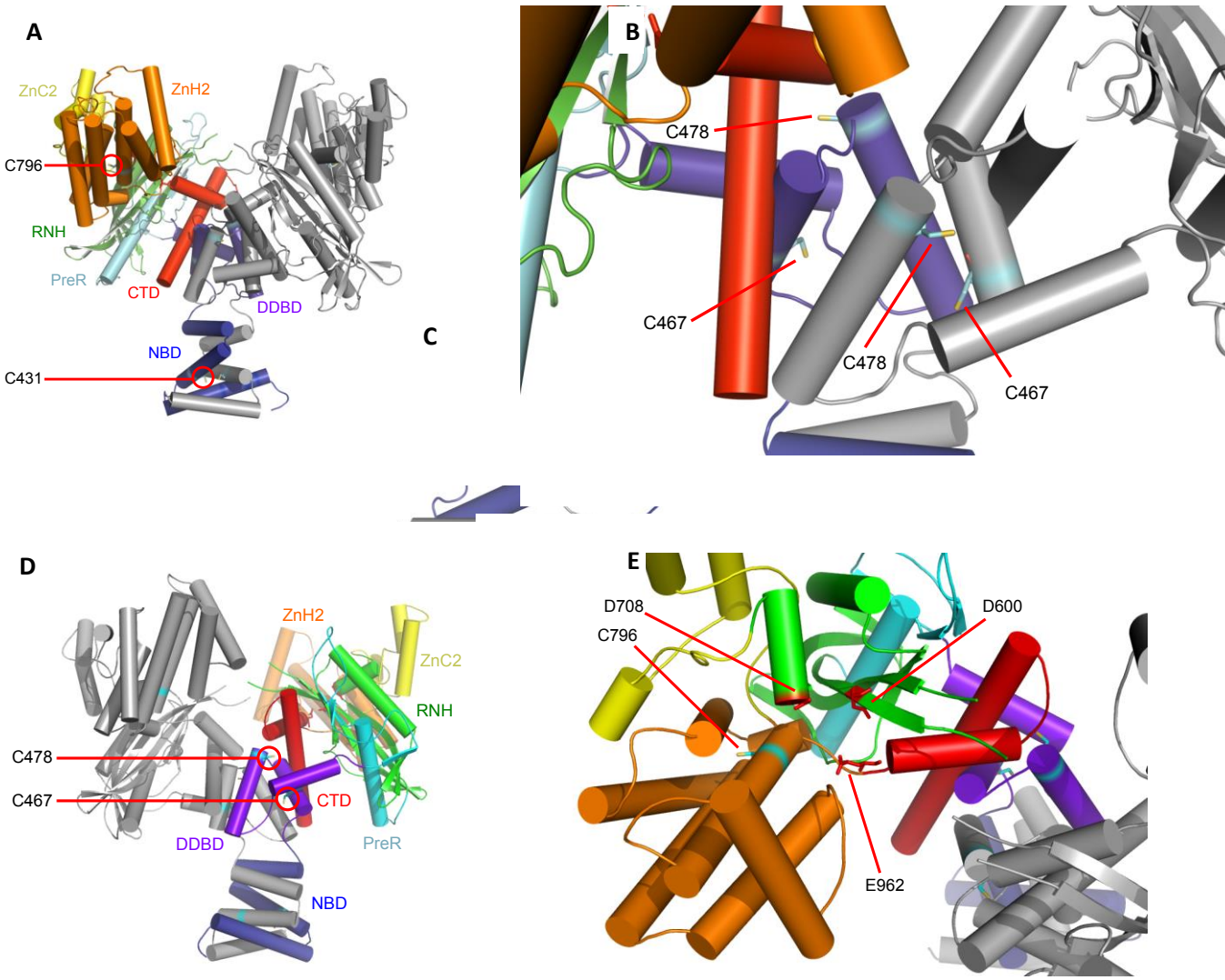


Figure 3.11 Localization of H3K4me3-dependent conformational changes to sites of DNA binding and catalysis in RAG-1. A. Locations of C431 and C476 in the NBD and ZnH2 domain, respectively, of RAG-1. The structure (Kim et al., 2015) was depicted from coordinates deposited in Protein Data Bank under accession code 4WWX. Only the RAG-1 chains are shown. Domains are labeled as defined in Fig. 4B. B. The RAG-1 dimer in (A) is rotated horizontally approximately 180˚ to display the locations of C467 and C478 in the DDBD. Domains are labeled as defined in Fig. 4B. C. H3K4me3-induced increase in solvent accessibility at C431, which lies at the NBD dimerization interface. D. H3K4me3-induced increases in solvent accessibility within the RAG-1 DDBD. C467 and C478 lie in two helical bundles that comprise a portion of the RSS heptamer binding site. E. The RAG-1 catalytic center. The residues that comprise the catalytic motif - D600, D708, and E962 - are highlighted in red. C796 lies within one of the helices of the ZnH2 domain, which bridges E962 to the remainder of the catalytic center.

cleavage (Lu 2015). In this regard it has been suggested (Liang 2002) that the RAG-2 non-core region stabilizes the association of RAG-1 with RAG-2 through an interaction with the DDBD of RAG-1. A speculative notion, which remains to be tested, is that the acidic autoinhibitory domain enjoys an association with the DDBD through charge-charge interactions, impairing RSS binding until inhibition is relieved by H3K4me3 Here we have favored the conservative interpretation that H3K4me3-induced changes in solvent accessibility represent changes in RAG conformation induced by occupancy of the PHD finger. A narrower interpretation is that the observed changes in solvent accessibility reflect exposure of residues that are occluded by the PHD finger in the absence of H3K4me3. While this interpretation is consistent with the available data, it seems unlikely that that all of these solvent accessibility changes occur at sites of direct contact with the PHD finger, because they occur in three separate regions of the RAG-1 surface – the NBD, the DDBD and the ZnH2 domain.

Allosteric regulation by epigenetic marks

Other enzymes in addition to RAG are subject to control by chromatin modifications. The histone methyltransferase polycomb repressive complex (PRC)2 is activated by binding of H3K27me3 to a WD40 propeller in the EED subunit (Byrum 2015), while the histone acetyltransferase Tip60 is stimulated by an interaction between its chromodomain and H3K9me3 (Margueron 2009). The DNA methyltransferase Dnmt3a, like RAG, is activated through a PHD finger, although in the case of Dnmt3a the allosteric ligand is H3K4me0 (Sun 2009). In none of these instances is it known how information regarding occupancy of the allosteric regulatory site is propagated within the protein. In Dnmt3a, mutations that impair allosteric activation but preserve binding to

H3K4me0 have been identified at a putative interface between the PHD and catalytic domains (Sun 2009), suggesting that propagation of the allosteric signal involves a direct interaction between these regions.

Evidence presented here indicates that a signal initiated by binding of H3K4me3 to the RAG-2 PHD is propagated to the RAG-2 autoinhibitory domain as well as to the DNA-binding and catalytic regions of RAG-1, but the pathways by which this signal is transmitted is unknown. Specifically, it is unclear whether allosteric signal transmission involves direct contact between the RAG-2 PHD and the catalytic or DNA-binding domains of RAG-1. Because available structures of RAG do not include the PHD or autoinhibitory regions of RAG-2 (Kim 2015, Ru 2015), it is not yet known where these domains lie relative to the sites of H3K4me3-induced conformational change in RAG-1.

Implications for coupling of V(D)J recombination to transcription

Initiation of V(D)J recombination is coupled to the lineage- and developmental stage-specific activation of transcription at specific antigen receptor loci. This coupling underlies the ordered rearrangement of heavy- and light-chain genes during lymphocyte development and provides a means by which Ig and TCR expression are restricted to the B- and T-lymphoid lineages, respectively. Transcriptional activity at antigen receptor loci is positively correlated with the local deposition of H3K4me3 (Li 2011), and the ability of H3K4me3 to serve as an allosteric activator of RAG suggests a molecular mechanism by which transcription and recombination are coupled. H3K4me3 in itself is insufficient, however, to establish locus specificity, as this epigenetic mark is enriched near all active promoters (Perkins 2004). Indeed, the distribution of RAG-2 in chromatin is positively correlated with the density of H3K4me3 (Barski 2007). The required

specificity could, in principle, be provided by combined recognition of H3K4me3 and one or more RSSs, although the participation of an additional ligand distinct from H3K4me3 remains possible. Results presented here suggest a sequential process of site specification in which the binding of H3K4me3 to the RAG-2 PHD induces conformational changes in the DNA-binding domains of RAG-1 that confer increased affinity for RSSs. The RAG heterotetramer contains two subunits each of RAG-1 and RAG-2. It is not known whether both PHD fingers must be occupied by H3K4me3 to effect allosteric activation, nor is it known whether the stimulatory effects of H3K4me3 on RSS binding and catalysis are exerted in cis or in trans, with respect to RAG-1. Answers to these questions will be critical to an understanding of how synapse formation and cleavage occur in the context of active chromatin.

Off-target RAG activity at cryptic cleavage sites is restricted to chromatin loop domains containing paired, canonical RSSs, and these off-target events are directionally oriented relative to the RSS ends (Hu 2015). These observations have suggested a directional sampling mechanism in which one canonical RSS in a paired complex dissociates from active RAG, allowing capture of a downstream cryptic RSS. While the mechanism underlying directional sampling by RAG is unknown, theoretical and experimental considerations (Hu 2015, Marcovitz 2013) suggest that RAG may assume distinct conformations in sampling mode and in its various RSS binding modes. If so, this suggests that H3K4me3 may modulate the transition from tracking to RSS capture by inducing a conformational shift in RAG that increases affinity for the RSS.

Experimental Methods

Pulse alkylation

Pulse alkylation was based on a published procedure (Chen 2012). RAG (0.5 μg active heterotetramer as determined by burst kinetic analysis) was incubated in in buffer R [25 mM HEPES (pH 7.4), 150 mM KCl, 0.5 mM DTT, 1 mM $CaCl_2$ and 10% glycerol] and 20 μM H3K4me3 or 20 μM H3K4me0 for 30 min at 37°C. Monobromobimane (mBBr) was added to 1mM. After 1 min at 37°C, alkylation was quenched by the addition of 10 mM β-mercaptoethanol and incubation was continued for 2 min to permit scavenging of unreacted mBBr. After quenching, protein was denatured by addition of guanidinium chloride (GdmCl) to 6M. Hexadeuterated mBBr (mBBr-d6) was added to 20 mM and samples were incubated for 2 hr at 37°C to permit alkylation of unreacted cysteines with mBBr-d6. Protein was dialyzed into 20 mM HEPES, 20% ACN, 5 mM β-mercaptoethanol, isolated by SDS page gel electrophoresis and visualized by staining with Coomassie blue. RAG-1 and RAG-2 were excised and in-gel digestion was carried out with trypsin and V8 protease in 50mM ammonium bicarbonate and 1% surfactant (ProteaseMax, Promega) as described (Shevchenko 2006). Peptides were extracted with 2:1 acetonitrile: 5% formic acid, lyophilized, resuspended in 0.1% trifluoroacetic acid and concentrated by C18 adsorption chromatography. Peptide samples were reconstituted in 10% formic acid and analysed by liquid chromatography-mass spectrometry.

Mass spectrometry data analysis and quantification

MS/MS spectra were searched against a customized database composed of wild-type RAG-1, wild-type RAG-2, RAG-2(W453A) and 248 common contaminant proteins using the SEQUEST search algorithm (Proteome Discoverer 1.4, Thermo Scientific). Cleavage by either trypsin or endoproteinase *Glu-C* was examined; a maximum of six missed cleavages were allowed. Alkylation of cysteine by mBBr or mBBr-d6 and oxidation of methionine were included as dynamic modifications. The minimum peptide length was set at 6 residues. The mass tolerances were set to 10 ppm. and 0.02 Da for precursor and fragment ions, respectively. Matched spectra were filtered using the Percolator algorithm (Proteome Discoverer). Peptide identifications were called at a q value <0.01.

Cysteine alkylation by mBBr and mBBr-d6 was quantified using a custom program coded in R (http://www.R-project.org/). Peptides were binned first according to the presence of cysteine and then according to the position of the cysteine in the RAG-1 or RAG-2 protein sequence. For each cysteine residue in RAG-1 or RAG-2, the total peak area, L_C or H_C, for all light-modified or heavy modified peptides was obtained by summation over all ^{13}C isotopologues and all sequential liquid chromatography samples in which the peptide was detected:

$$\sum_{p=0}^{n} C(p, L) \equiv L_C \quad \text{and} \quad \sum_{p=0}^{n} C(p, H) \equiv H_C,$$

where C is the peptide corresponding to a specific cysteine residue in the RAG-1 or RAG-2 sequence, p denotes an individual MS peak corresponding to that residue, L_C denotes alkylation of residue C by mBBr and H_C denotes alkylation of residue C by mBBr-d6. The relative accessibility of each modified cysteine residue to light mBBR,

which is interpreted to indicate solvent accessibility in one or more conformational states, was determined by:

$$L_C/(L_C + H_C).$$

The change in accessibility of a particular cysteine residue associated with the introduction of the plant homeodomain ligand H3K4me3 is then:

$$[L_C/(L_C + H_C)]_{me3} - [L_C/(L_C + H_C)]_{me0},$$

where the subscripts me3 and me0 specify relative accessibility in the presence of H3K4me3 and H3K4me0, respectively.

Cell culture

HEK 293T cells were grown in Dulbecco's Modified Eagle Medium (DMEM) lacking pyruvate and supplemented with 10% fetal bovine serum and 1X penicillin-streptomycin-glutamine (Life Technologies). Cells were maintained at 37°C in 5% CO_2.

Expression constructs

DNA fragments encoding full length wild-type RAG-2 and RAG-2(W453A) were PCR amplified and ligated into pcDNA3.1, in frame with a cassette encoding an amino-terminal MBP tag and a carboxy-terminal tobacco etch virus protease site (MBP-TEV). To generate an MBP tagged form of cR1ct, a DNA fragment encoding amino acid residues 384-1040 of RAG-1 was cloned into pcDNA 3.1 in frame with an amino terminal MBP-TEV cassette.

Purification of RAG proteins

Maltose binding protein (MBP) fusions to full length wild-type RAG-2, RAG-2(W453A), or amino acid residues 383-1040 of RAG-1 (cRAG1-ct) were coexpressed in 293T cells and purified by amylose affinity chromatography as described (Lu et al., 2015). The active fraction of each RAG protein preparation was determined by burst kinetic analysis as described (Lu et al., 2015). Equivalent molar amounts of active RAG, as determined by burst kinetic analysis, were compared in coupled DNA cleavage and pulse alkylation mass spectrometry assays.

Oligonucleotide substrates

The double-stranded oligonucleotide substrates HL-44/45 (12-RSS) and HL-46/47 (23-RSS) were prepared as described (Shimazaki et al., 2012). The 5' end of oligonucleotide HL-44 was radiolabeled with ^{32}P by T4 DNA polynucleotide kinase (New England Biolabs) before annealing to HL-45. Duplex oligonucleotides were purified by gel electrophoresis.

Assays for coupled DNA cleavage

Coupled cleavage assays were performed as described (Lu et al., 2015). Equivalent amounts of active wild-type or mutant RAG tetramer (1 nM), as determined by burst kinetic analysis, were incubated with 5 nM ^{32}P-labeled HL44/45 substrate, 5 nM unlabeled HL46/47 substrate and varying amounts of H3K4me3 or H3K4me0 peptide in 10 μL reaction buffer (25 mM 3-(4-Mopholino)propane sulfonic acid-KOH pH 7.0, 30mM KCl, 30 mM potassium glutamate, 5 mM $CaCl_2$, 0.5 mM DTT, 0.1 mg/mL BSA, 1% glycerol). After 20 min at 37°C, $MgCl_2$ was added to 5 mM and incubation was continued for 1 hr. Reactions were stopped by addition of loading buffer (90%

formamide, 0.2% bromophenol blue) and incubation at 90°C for 2 min. Products were seperated on a 15% polyacrylamide-urea gel, analyzed on a phosphor imager, and quantified using ImageQuantNL.

Liquid chromatography-mass spectrometry

Peptide samples were reconstituted in 10% formic acid and analyzed by mass spectrometry (MS) on a linear trap quadrupole (LTQ)-Orbitrap Elite mass spectrometer (Thermo Electron), coupled to a nanoflow liquid chromatography system (Easy-nLCII, Thermo Scientific). Peptides were resolved on a C_{18} analytical column (75 □m × 60 cm, 5 μm particle size) at a flow rate of 250 nl min^{-1} using a linear gradient of 7% – 35% solvent B (0.1% formic acid, 95% acetonitrile) over 65 min. Spectra were acquired with full scans (300 – 1,700 *m/z*) using the Orbitrap mass analyzer at a mass resolution of 120,000 at 400 *m/z*. The ten most intense precursor ions from a survey scan were selected for tandem mass spectrometry (MS/MS) from each duty cycle. The selected precursor ions were fragmented by higher-energy collision dissociation and detected at a mass resolution of 30,000 at 400 *m/z*. Dynamic exclusion was set for 30 sec with a 10 ppm mass window. The automatic gain control was set to 1 million for full Fourier transform (FT) MS and to 0.05 million for FT MS/MS, with maximum ion injection times of 100 ms and 300 ms, respectively. Internal calibration was carried out using lock mass from ambient air (*m/z* 445.1200025). Mass spectroscopic images were generated using mMass (Niedermeyer and Strohalm, 2012).

Limited proteolysis

RAG was preincubated with H3K4me3 or H3K4me0 in buffer R as described for pulse alkylation, except that $CaCl_2$ was added to 0.5mM. Thermolysin (Promega) was added at a protein mass ratio of 1:20 thermolysin:RAG and incubation was continued at 37°C for 15 min. Samples were removed at 5, 10, and 15 min and reactions were stopped by heating to 100°C in SDS PAGE loading buffer containing 50 mM EDTA. Products were fractionated by 9% SDS PAGE, detected by staining with Coomassie blue, excised and digested with either trypsin (Promega) or Lys-C (Promega). Peptides were analyzed by LC-MS/MS as described above for PA-MS. Thermolysin cleavage sites were determined by first comparing peptides from the thermolysin partial proteolysis band to control RAG-1 and RAG-2 proteins digested with the secondary protease (trypsin or Lys-C), and identifying unique peptides present in only the partial proteolysis band, that begin or terminate in an N-terminal cleavage to one of amino acids: F,V,L,I,A, or M.

Accessible surface area calculations

Accessible surface areas of cysteine residues in RAG were calculated using the Protein interfaces, surfaces and assemblies (PISA) of the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html), using available structures of the RAG tetramer (PDB accession codes 4WWX and 3JBW).

Supplement: Identifying the RAG-1/RAG-2 allosteric interface through least-common ancestor substitution.

Results

From our pulse alkylation studies of the RAG recombinase, we predicted that mutations could be generated which would permit RAG-2 to bind to H3K4me3, but would abolish transmission of this allosteric signal to RAG-1. To identify such mutations, and to interrogate the path through which allostery is transmitted from the RAG-2 PHD finger, we performed a blast alignment of RAG-2 proteins from all species in the NCBI database (version 2.6.0), to find the furthest divergent RAG-2 PHD finger sequences from Mus *musculus*. Of those identified, we selected two sequences from the shark

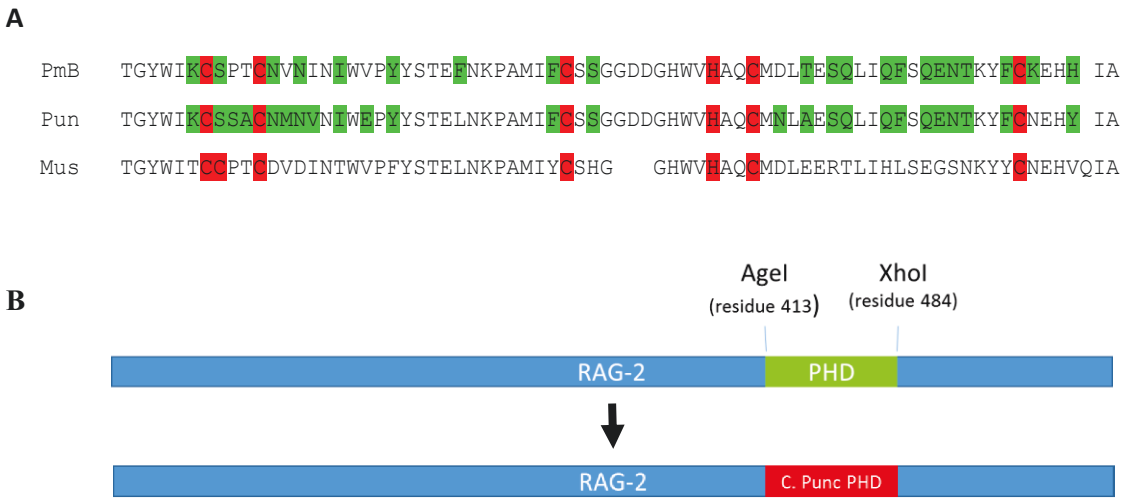


Figure 4.1. Generation of RAG-2 PHD finger chimeras. A. BLAST Alignment of the M. *musculus* (Mus) and Shark PHD fingers. C. *plumbeus* (Pmb), C. *punctatum* (Pun). Non-conserved residues in the shark PHDs are highlighted in green. The highly conserved Cys-His-Cys motif is highlighted in red. B. Strategy for inserting the shark PHD fingers into murine RAG-2. Silent mutations were generated in a MBP tagged RAG-2 cDNA clone which allowed restriction digest excision of the murine plant homeodomain. After digestion, the shark PHD fingers were ligated in frame into murine RAG-2, producing the desired chimeric proteins.

Chiloscyllium *punctatum* and Carcharhinus *plumbeus*, and substituted these PHD fingers for the PHD finger of mouse RAG-2. The chimeric shark PHD RAG-2 proteins contain mouse sequence from residues 1-413 and from 485-512; the shark PHD finger sequence

extends from 414 to 484 (Fig 4.1 A,B). To determine whether substitution of either the C. punctatum or C. plumbeus PHD finger disrupts RAG-2 mediated activation of RAG-1, we employed the PJH200 assay, which is an extrachromosomal assay for V(D)J recombination. Recombination was impaired by the substitution of the C. punctatum PHD finger into murine RAG-2, but was not significantly impaired by C. plumbeus (Fig 4.2A). Mutation of the critical tryptophan W456 within the PHD pocket, which ablates PHD-H3K4me3 interactions, did not significantly decrease the activity of C. punctatum Rag-2, but did decrease the activity of C. plumbeus RAG-2.

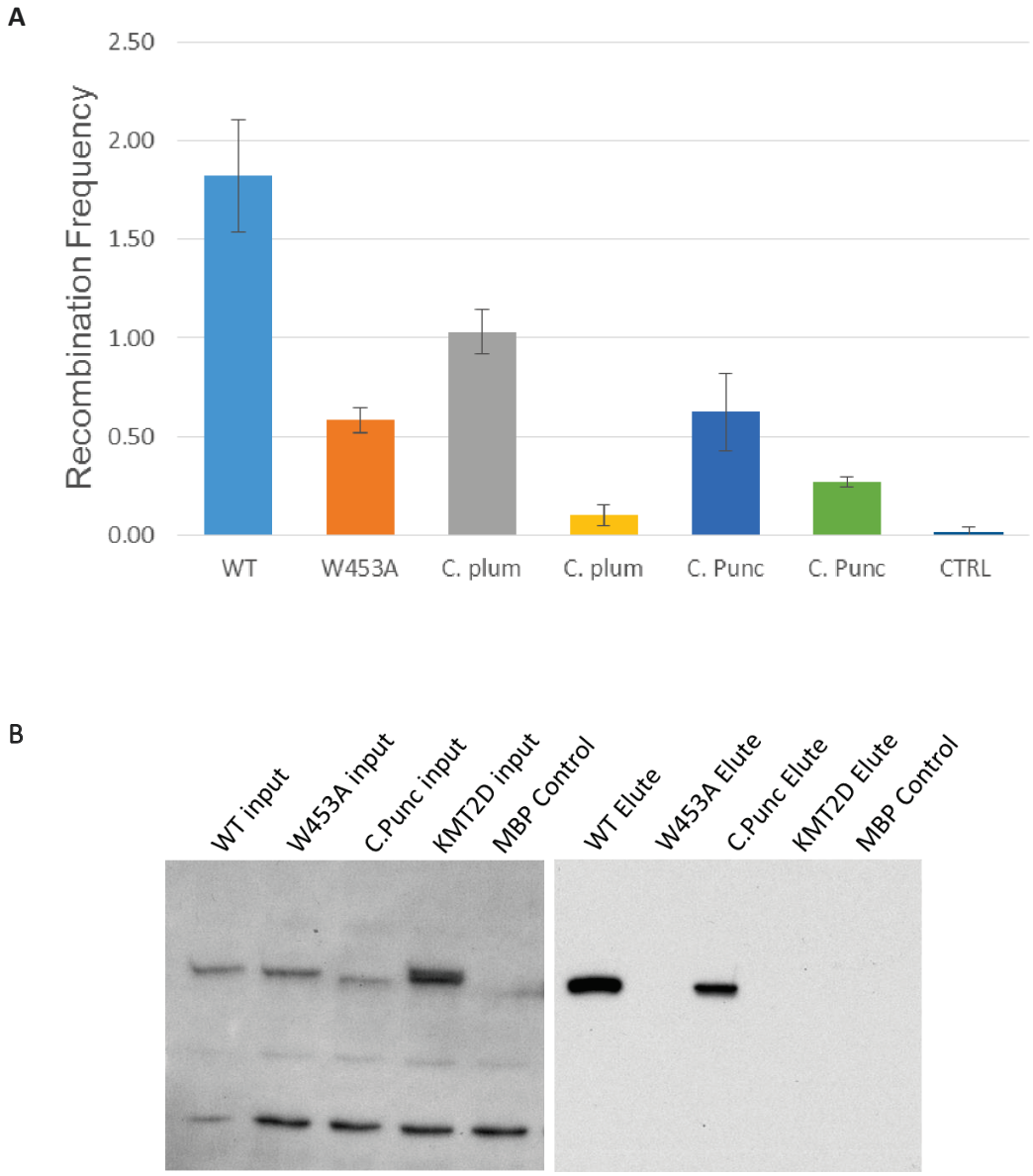


Figure 4.2 Effect of Shark PHD finger substitution in murine RAG-2. A. Effect of C. Plumbeus and C. Punctatum PHD/murine RAG-2 chimeras on V(D)J recombination. B. Pulldown assay of WT, W453A, and C. Punctatum RAG-2 with H3K4me3 beads. 1% of input is shown in the left panel, and elution fractions from the beads after washing is shown on the right. Kmt2D was an additional chimera tested that is unrelated to the shark PHD chimeras, which did not exhibit H3K4me3 binding and was discarded.

In order to verify that the C. punctatum chimeric RAG-2 was still capable of binding H3K4me3, we expressed MPB tagged versions of WT, W453A, and C. punctatum RAG-2 proteins in HEk-293T cells, and performed co-immunoprecipitations to determine whether these RAG-2 proteins bound H3K4me3. H3K4me3 co-precipitated with both WT and C. Punctatum RAG-2, but not W453A RAG-2, indicating that the C. punctatum RAG-2 chimera was still competent to bind H3K4me3 (Fig 4.2B).

The C. punctatum RAG-2 chimera exhibited a loss of function *in vivo*, as demonstrated by the pJH200 assay. However, this loss of function could have been due to instability of the chimeric protein because of the large PHD substitution. To confirm that this protein was a true allosteric null mutant of RAG-2, and not a non-functional or

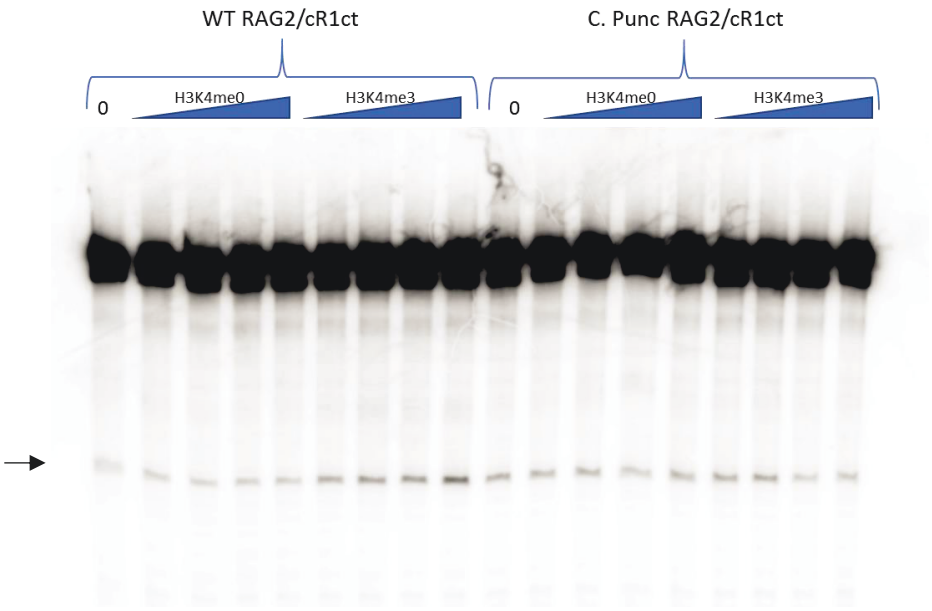


Figure 4.3 H3K4me3 stimulates DNA cleavage by RAG complexes containing wild-type RAG-2 but not C. *punctatum* RAG-2. Coupled cleavage reactions contained radiolabeled 12-RSS and unlabeled 23-RSS. Reactions contained wild-type (WT) RAG-2 or C. *punctatum* RAG-2 as defined at top. K4me0 and K4me3, reactions supplemented with 0.5, 1, 2 or 4 μM H3K4me0 or H3K4me3 peptide; 0, reactions lacking peptide. Arrow indicates nicked DNA product produced by RAG cutting.

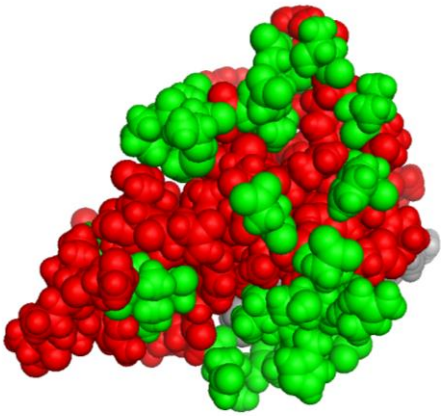
misfolded protein, we purified C. punctatum and WT RAG-2 with cRAG1-CT, and tested them in a DNA cleavage assay. Both RAG-2 proteins were stimulated either with control H3K4me0 peptide, or H3K4me3, to determine if they were responsive to H3K4me3. As previously observed, the cleavage activity of WT RAG-2 in complex with cRAG-1-CT exhibited a dose dependent increase in DNA cleavage in response to H3K4me3 stimulation (Fig 4.3). In contrast, the C. *punctatum* RAG-2 chimera exhibited a basal level of cleavage similar to that of WT RAG-2, but was unresponsive by H3K4me3.

The C. punctatum RAG-2 chimera is competent to bind H3K4me3, exhibits a loss of function *in vivo*, and is active but unstimulable by H3K4me3 in an *in vitro* DNA cleavage assay, meeting the criteria for an allosterically null mutant of RAG-2. We predicted therefore that one or more of the nonconserved residues of the C. punctatum PHD which had been substituted into murine RAG-2 had disrupted the allosteric interface between RAG-1 and RAG-2. This suggests that reverting one or more of the substituted residues back to native murine sequence should rescue loss of function phenotype. We tested this hypothesis by back-mutating groups of 5-6 amino acids in the C. punctatum PHD chimera to the WT mouse sequence (Fig 4.4).

A

```
M. Mus   TGYWITCCPTCDVDINTWVPFYSTELNKPAMIYCSHG---GHWVHAQCMDLEERTLIHLSEGSNKYYCNEHVQIA
C. Punc  TGYWIKCSSACNMNVNIWEPYYSTELNKPAMIFCSSGGDDGHWVHAQCMNLAESQLIQFSQENTKYFCNEHYIAR
Group-1  TGYWITCCPTCDMNVNIWEPYYSTELNKPAMIFCSSGGDDGHWVHAQCMNLAESQLIQFSQENTKYFCNEHYIAR
Group-1  TGYWIKCSSACNVDINTWVPFYSTELNKPAMIFCSSGGDDGHWVHAQCMNLAESQLIQFSQENTKYFCNEHYIAR
Group-3  TGYWIKCSSACNMNVNIWEPYYSTELNKPAMIFCSSGGDDGHWVHAQCMDLEERTLIHLSQENTKYFCNEHYIAR
Group-4  TGYWIKCSSACNMNVNIWEPYYSTELNKPAMIFCSSGGDDGHWVHAQCMNLAESQLIQFSEGSNKYFCNEHYIAR
Group-5  TGYWIKCSSACNMNVNIWEPYYSTELNKPAMIYCSHGGDDGHWVHAQCMNLAESQLIQFSQENTKYYCNEHVIAR
```

B



C

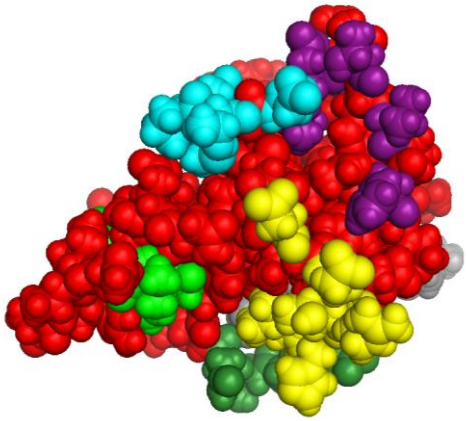


Figure 4.4 Reversion mutation of the C. *punctatum* RAG-2 chimera back to M. *musculus* sequence by group substitution. A. Sequence alignments of the M. *musculus* and C. *punctatum* PHD fingers, and the reversion mutants with shark-to-mouse back mutations highlighted. B. Top down view of the crystal structure of the RAG-2 PHD finger, with non-synonomous mutations between C. *punctatum* and mouse highlighted in green. Red residues are conserved between mouse and shark, H3K4me3 is highlighted in Gray. C. Mapping of the reversion mutant groups in the PHD finger. The green residues are now highlighted by reversion group color: Group 1- Blue, Group 2- Purple, Group 3- Yellow, Group 4- Teal, Group 5- Green. Note that although Group 5 substitutions are broadly distributed in primary sequence, they are tightly associated in the 3 dimensional structure.

These back-mutated RAG-2 variants were then tested in the recombination assay, to determine whether any of these reversion mutations rescued the C. punctatum phenotype (Fig 4.5 A). In some repeats of this experiment, group 1 and group 2 reversion mutants both exhibited a partial restoration of RAG function; this suggested the interface

that transmits the allosteric signal from RAG-2 to RAG-1 may be bounded by both these groups of residues. We therefore generated a revertant mutant in which both group 1 and group 2 were back-mutated to murine sequence, and then assayed this variant for recombination. Reversion of both group 1 and group 2 residues rescued RAG activity to a level near that of WT, a rescue that was better than either group 1 or group 2 (Fig 4.5 B). These results suggest that the allosteric interface of RAG-2 is bounded between residues 417 and 433.

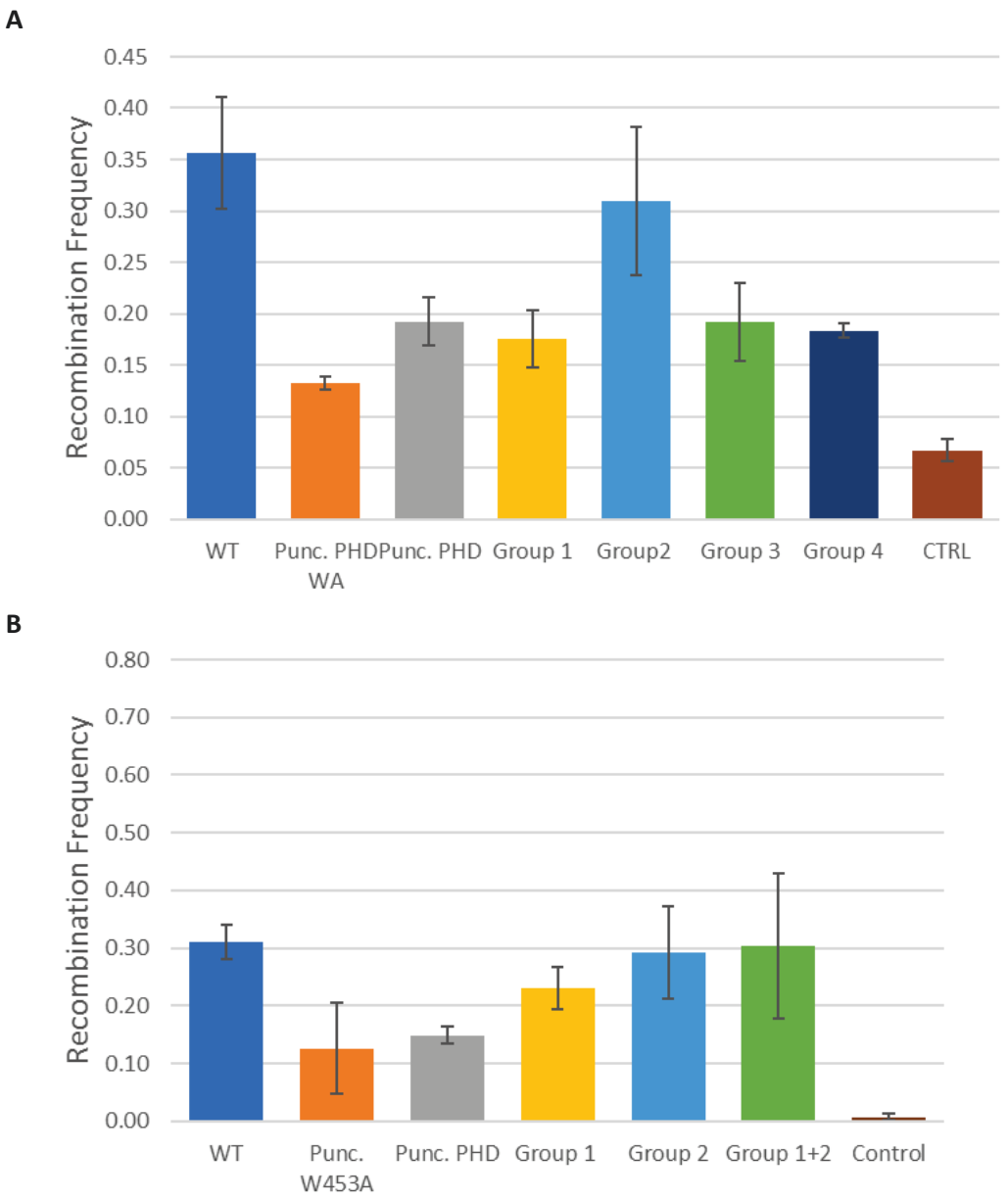


Figure 4.5 Rescue of the inactivating C. *punctatum* substitution by group back mutation. A. PJH200 assay utilizing the PJH200 substrate, full-length RAG-1, and the indicated RAG-2 variant: Control (empty vector), wild-type RAG-2 (WT), or indicated mutant) were transfected into NIH3T3 cells. The bars represent the mean ± SEM, n = 3 independent transfections. Recombination (%) is a calculation of the frequency of signal joining of the pJH200 signal joint recombination substrate. B. The same assay as in A. including a new RAG mutant containing both Group 1 and Group 2 reversion mutants in a single RAG-2 construct.

General Discussion

Antigen receptors are a canonical component of the mammalian immune system; these unique proteins recognize and target for elimination the innumerable foreign invaders that attack the body every day. Underlying these proteins is V(D)J recombination, initiated by the RAG recombinase, which generates the primary diversity of the antigen receptor genes. Accessibility of antigen receptor genes to RAG is transcription dependent, and the presence of "active" chomatin marks such as H3K4 trimethylation are positively correlated with the ability to undergo recombination. This accessibility hypothesis initially posited that the "openness" of the antigen receptor loci chromatin loci was the key element in driving recombination (Yancopoulos 1985). Our findings refine this observation, and provide a new model for regulation of nuclear enzymes by chromatin. The observation that the W453A nearly ablates recombination demonstrates that transcription of an antigen receptor locus is not sufficient to drive recombination, Rather, recombination requires coordination of locus accessibility with specific chromatin contacts with RAG that promote V(D)J recombination. Our *in vitro* characterizations of the effects of H3K4me3 on RAG activity indicate that engagement of the RAG-2 PHD finger with H3K4me3 has profound stimulatory effects on both the affinity of RAG for and RSS and DNA cleavage.

Pulse alkylation studies carried out here provide the first direct evidence to explain these H3K4me3 changes in Kcat and K_D. H3K4me3 stimulation of RAG is associated with increased solvent accessibility of the DNA binding domains of RAG-1, as well as the catalytic center. While not detected by pulse alkylation, concomitant increases in solvent exposure of the RAG-2 inhibitory domain was observed by partial proteolysis. These changes in solvent accessibility likely represent allosteric transitions

within the RAG complex. Allosteric regulation of proteins is a well-known phenomenon; allostery gates the activity of many important enzymes including members of kinase cascades, integrases, and a number of oncogenes. But epigenetic allosteric regulation of nuclear enzymes to this point have been largely unexplored

Three lines of evidence support the interpretation that H3K4me3 is an allosteric regulator of RAG. First, an H3K4me3 peptide stimulates RAG in vitro, and in the absence of additional cofactors. Second, pulse alkylation mass spectrometry footprinting of the RAG proteins in the presence of H3K4me3 reveals that several domains of RAG-1 exhibit a conformational changes upon H3K4me3 binding to the RAG-2 PHD finger. Third, RAG-2 mutants have been generated that are capable of binding H3K4me3 but unable to stimulate RAG-1 catalytic activity. These data are consistent with a model in which the inhibitory domain of RAG-2 occupies the DNA binding domain(s) of RAG-1, until such time that RAG-2 engages the H3K4me3 mark. In this model, engagement of RAG-2 with H3K4me3 would relieve inhibition of RAG-1, permitting enhanced DNA binding and cleavage.

` Although an allosteric change was detected in the RAG-2 AcID domain (Ward unpublished data) using limited proteolysis, this domain does not appear to be responsible for transmitting the allosteric binding signal from RAG-2 to RAG-1. Our mapping of the RAG-2 allosteric transmission site by least-common ancestor substitution indicates that non-synonomous mutations in a PHD finger from the C. punctatum shark RAG-2, when placed in the context of murine RAG-2, disrupts the allosteric transmission signal generated by binding H3K4me3. The C. punctatum RAG-2 PHD finger murine RAG-2 chimera was unable to support V(D)J recombination on an extrachromosomal

substrate (PJH200) but retained the ability to engage the H3K4me3 mark. This mutant was the first evidence that the ability of RAG-2 to bind H3K4me3 is not sufficient to stimulate RAG activity, and that localization of RAG-2 with H3K4me3 in the nucleus is not the primary function of the RAG-2 PHD finger. Selective reversion of non-synonomous mutations within the C. punctatum PHD back to murine sequence rescued this loss of function phenotype, indicating that residues 417-433 of RAG-2 are responsible for transmission of the allosteric signal from RAG-2 to RAG-1. Structural mapping of these mutations make physiological sense, as these residues form an interface on the exterior of the RAG-2 PHD finger, directly above the H3K4me3 binding pocket. These data strongly argue against H3K4me3 as a mere localization signal for RAG-2. Instead, the critical role of the PHD finger appears to be to impart these allosteric changes to the RAG complex as a whole upon binding H3k4me3. These data further reinforce the model that H3K4me3 allosterically regulates RAG through the RAG-2 PHD finger.

Parallel work to this thesis by Ward and others (Liu 2015 and unpublished data) revealed the existence of an "AcID" (acidic inhibitory domain) region within RAG-2 that when mutated, can bypass RAG's dependence on H3K4me3, rescuing the W453A phenotype. This initially suggested that the activity of these domains might be linked. However, the AcID domain can act independently of the PHD finger, although the two domains do appear to have direct contacts (Ward unpublished data). Still, the exact nature of the interplay between the AcID domain and the PHD finger of RAG-2 warrant further study.

Allosteric regulation of proteins by chromatin may be a more general phenomenon of nuclear enzyme complexes. Many nuclear enzyme complexes contain a

subunit capable of recognizing one or more post translational modifications. Allosteric gating of these enzyme complexes would provide a means to suppress activity, until such a time when they encounter their allosteric stimulating chromatin mark, thereby limiting off target activity. While this hypothesis remains to be tested, it is interesting to speculate that the observations made herein regarding RAG may be generalizable to other nuclear proteins.

Supplementary Experimental Methods:

RAG-2 Pulldowns

10 μg MBP Shark chimera RAG-2 (Either C. punctatum or C. plumbeus) were transfected into HEK-293T cells plated at $2*10^6$ cells/plate, using the PEI. 48 hours post transfection cells were harvested in ice cold PBS containing 2mM EDTA, and cells were pelleted at 500Xg for 5 minutes. Pellets were resuspended in 500μL lysis buffer (50mMTris (pH 7.5), 300mMNaCl, 1 mMPMSF,1%NP-40, 1%deoxycholic acid, 0.1% SDS, and protease inhibitor cocktail (Roche)). Lysates were clarified by centrifugation at 20,000XG for 30 minutes at 4°C. Clarified lysates were diluted 10-fold in binding buffer (50 mM Tris [pH 7.5], 300 mM NaCl, 1 mM PMSF, and protease inhibitors) then with 100μL of 5 mg H3K4me3 peptide prebound to streptavidin agarose for 4 hr at 4°C. Beads were washed three times with binding buffer supplemented with 0.1% NP-40. Proteins were eluted in boiling SDS loading buffer, then run on a 9% SDS Page gel, and analyzed by western blot.

PJH200 Assays

10 μg each of MBP-RAG-1-myc-his (full length) and MBP-RAG-2 (wild-type or mutant, as indicated in the figures above), and 5 μg of pJH200 (Hesse et al. 1987). Plasmids were co-transfected into NIH3T3 cells in a 10 cm dish with TransIT-LT1 (Mirus, Cat#MIR 2300) or Lipofectamine 20000 (Thermo Fisher, Cat#11668-019) per manufacturer's protocol. Recombination efficiency was similar with both protocols (results not shown). After 48 hours, the cell pellet was divided into thirds, one for storage, one for DNA isolation, and one for protein extraction. Plasmid DNA was

extracted by a modified Hirt extraction per manufacturer's protocol (Qiagen). 50 μg isolated DNA was transformed into 50 μL DH5α Max Efficiency cells (Thermo Fisher, Cat#18258012) per manufacturer's protocol. 1.7% (5μL) of the transformation mixture was plated on LB agar containing 50 μg / mL ampicillin and the rest was plated on LB agar containing 50 μg / mL ampicillin and 15 μg / mL chloramphenicol. Plates containing ampicillin alone were scored after 16 hours at 37 °C, while plates with ampicillin and chloramphenicol were scored after 18 hours.

Assays for DNA cleavage

Coupled cleavage assays were performed as described (Lu et al., 2015). Equivalent amounts of active wild-type or mutant RAG tetramer (5μg) were incubated with 5 nM ^{32}P-labeled HL44/45 substrate, 5 nM unlabeled HL46/47 substrate and varying amounts of H3K4me3 or H3K4me0 peptide in 10 μL reaction buffer (25 mM 3-(4-Mopholino)propane sulfonic acid-KOH pH 7.0, 30mM KCl, 30 mM potassium glutamate, 5 mM $CaCl_2$, 0.5 mM DTT, 0.1 mg/mL BSA, 1% glycerol). After 20 min at 37°C, $MgCl_2$ was added to 5 mM and incubation was continued for 1 hr. Reactions were stopped by addition of loading buffer (90% formamide, 0.2% bromophenol blue) and incubation at 90°C for 2 min. Products were separated on a 15% polyacrylamide-urea gel, analyzed on a phosphor imager, and quantified using ImageQuantNL.

References

Afshar R, Pierce S, Bolland DJ, Corcoran A, Oltz EM. Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination. J Immunol. 2006 Feb 15;176(4):2439-47.

Akamatsu Y, Monroe R, Dudley DD, Elkin SK, Gartner F, Talukder SR, et al. Deletion of the RAG2 C terminus leads to impaired lymphoid development in mice. Proc Natl Acad Sci U S A. 2003 Feb 4;100(3):1209-14.

Alt FW, Yancopoulos GD, Blackwell TK, Wood C, Thomas E, Boss M, et al. Ordered rearrangement of immunoglobulin heavy chain variable region segments. EMBO J. 1984 Jun;3(6):1209-19.

Bain G, Quong MW, Soloff RS, Hedrick SM, Murre C. Thymocyte maturation is regulated by the activity of the helix-loop-helix protein, E47. J Exp Med. 1999 Dec 6;190(11):1605-16.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007 May 18;129(4):823-37.

Baumann M, Mamais A, McBlane F, Xiao H, Boyes J. Regulation of V(D)J recombination by nucleosome positioning at recombination signal sequences. EMBO J. 2003 Oct 1;22(19):5197-207.

Bevington S, Boyes J. Transcription-coupled eviction of histones H2A/H2B governs V(D)J recombination. EMBO J. 2013 May 15;32(10):1381-92.

Biswas T, Aihara H, Radman-Livaja M, Filman D, Landy A, Ellenberger T. A structural basis for allosteric control of DNA recombination by lambda integrase. Nature. 2005 Jun 23;435(7045):1059-66.

Bolland DJ, Koohy H, Wood AL, Matheson LS, Krueger F, Stubbington MJ, et al. Two Mutually Exclusive Local Chromatin States Drive Efficient V(D)J Recombination. Cell Rep. 2016 Jun 14;15(11):2475-87.

Bolland DJ, Wood AL, Johnston CM, Bunting SF, Morgan G, Chakalova L, et al. Antisense intergenic transcription in V(D)J recombination. Nat Immunol. 2004 Jun;5(6):630-7.

Callebaut I, Mornon JP. The V(D)J recombination activating protein RAG2 consists of a six-bladed propeller and a PHD fingerlike domain, as revealed by sequence analysis. Cell Mol Life Sci. 1998 Aug;54(8):880-91.

Chakraborty T, Chowdhury D, Keyes A, Jani A, Subrahmanyam R, Ivanova I, et al. Repeat organization and epigenetic regulation of the DH-Cmu domain of the immunoglobulin heavy-chain gene locus. Mol Cell. 2007 Sep 7;27(5):842-50.

Chen YT, Collins TR, Guan Z, Chen VB, Hsieh TS. Probing conformational changes in human DNA topoisomerase IIalpha by pulsed alkylation mass spectrometry. J Biol Chem. 2012 Jul 20;287(30):25660-8.

Cobb RM, Oestreich KJ, Osipovich OA, Oltz EM. Accessibility control of V(D)J recombination. Adv Immunol. 2006;91:45-109.

Cuomo CA, Oettinger MA. Analysis of regions of RAG-2 important for V(D)J recombination. Nucleic Acids Res. 1994 May 25;22(10):1810-4.

Fuxa M, Skok J, Souabni A, Salvagiotto G, Roldan E, Busslinger M. Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. Genes Dev. 2004 Feb 15;18(4):411-22.

Gellert M. V(D)J recombination: RAG proteins, repair factors, and regulation. Annu Rev Biochem. 2002;71:101-32.

Gigi V, Lewis S, Shestova O, Mijuskovic M, Deriano L, Meng W, et al. RAG2 mutants alter DSB repair pathway choice in vivo and illuminate the nature of 'alternative NHEJ'. Nucleic Acids Res. 2014 Jun;42(10):6352-64.

Goldmit M, Ji Y, Skok J, Roldan E, Jung S, Cedar H, et al. Epigenetic ontogeny of the Igk locus during B cell development. Nat Immunol. 2005 Feb;6(2):198-203.

Grundy GJ, Yang W, Gellert M. Autoinhibition of DNA cleavage mediated by RAG1 and RAG2 is overcome by an epigenetic signal in V(D)J recombination. Proc Natl Acad Sci U S A. 2010 Dec 28;107(52):22487-92.

Guo C, Yoon HS, Franklin A, Jain S, Ebert A, Cheng HL, et al. CTCF-binding elements mediate control of V(D)J recombination. Nature. 2011 Sep 11;477(7365):424-30.

Guo F, Gopaul DN, van Duyne GD. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature. 1997 Sep 4;389(6646):40-6.

Guo F, Gopaul DN, van Duyne GD. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature. 1997 Sep 4;389(6646):40-6.

Hu J, Zhang Y, Zhao L, Frock RL, Du Z, Meyers RM, et al. Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. Cell. 2015 Nov 5;163(4):947-59.

Ji Y, Little AJ, Banerjee JK, Hao B, Oltz EM, Krangel MS, et al. Promoters, enhancers, and transcription target RAG1 binding during V(D)J recombination. J Exp Med. 2010 Dec 20;207(13):2809-16.

Jiang H, Chang FC, Ross AE, Lee J, Nakayama K, Nakayama K, et al. Ubiquitylation of RAG-2 by Skp2-SCF links destruction of the V(D)J recombinase to the cell cycle. Mol Cell. 2005 Jun 10;18(6):699-709.

Johnson K, Pflugh DL, Yu D, Hesslein DG, Lin KI, Bothwell AL, et al. B cell-specific loss of histone 3 lysine 9 methylation in the V(H) locus depends on Pax5. Nat Immunol. 2004 Aug;5(8):853-61.

Jung D, Giallourakis C, Mostoslavsky R, Alt FW. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. Annu Rev Immunol. 2006;24:541-70.

Kassabov SR, Zhang B, Persinger J, Bartholomew B. SWI/SNF unwraps, slides, and rewraps the nucleosome. Mol Cell. 2003 Feb;11(2):391-403.

Kim MS, Lapkouski M, Yang W, Gellert M. Crystal structure of the V(D)J recombinase RAG1-RAG2. Nature. 2015 Feb 26;518(7540):507-11.

Kirch SA, Rathbun GA, Oettinger MA. Dual role of RAG2 in V(D)J recombination: catalysis and regulation of ordered Ig gene assembly. EMBO J. 1998 Aug 17;17(16):4881-6.

Klemm RD, Austin RJ, Bell SP. Coordinate binding of ATP and origin DNA regulates the ATPase activity of the origin recognition complex. Cell. 1997 Feb 21;88(4):493-502.

Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG, et al. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science. 2002 Apr 5;296(5565):158-62.

Lee J, Desiderio S. Cyclin A/CDK2 regulates V(D)J recombination by coordinating RAG-2 accumulation and DNA repair. Immunity. 1999 Dec;11(6):771-81.

Li BZ, Huang Z, Cui QY, Song XH, Du L, Jeltsch A, et al. Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. Cell Res. 2011 Aug;21(8):1172-81.

Li Z, Dordai DI, Lee J, Desiderio S. A conserved degradation signal regulates RAG-2 accumulation during cell division and links V(D)J recombination to the cell cycle. Immunity. 1996 Dec;5(6):575-89.

Liang HE, Hsu LY, Cado D, Cowell LG, Kelsoe G, Schlissel MS. The "dispensable" portion of RAG2 is necessary for efficient V-to-DJ rearrangement during B and T cell development. Immunity. 2002 Nov;17(5):639-51.

Liu Y, Subrahmanyam R, Chakraborty T, Sen R, Desiderio S. A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. Immunity. 2007 Oct;27(4):561-71.

Lu C, Ward A, Bettridge J, Liu Y, Desiderio S. An autoregulatory mechanism imposes allosteric control on the V(D)J recombinase by histone H3 methylation. Cell Rep. 2015 Jan 6;10(1):29-38.

Maman Y, Teng G, Seth R, Kleinstein SH, Schatz DG. RAG1 targeting in the genome is dominated by chromatin interactions mediated by the non-core regions of RAG1 and RAG2. Nucleic Acids Res. 2016 Nov 16;44(20):9624-37.

Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ,3rd, et al. Role of the polycomb protein EED in the propagation of repressive histone marks. Nature. 2009 Oct 8;461(7265):762-7.

Marsh JA, Teichmann SA. Relative solvent accessible surface area predicts protein conformational changes upon binding. Structure. 2011 Jun 8;19(6):859-67.

Mather EL, Perry RP. Methylation status and DNase I sensitivity of immunoglobulin genes: changes associated with rearrangement. Proc Natl Acad Sci U S A. 1983 Aug;80(15):4689-93.

Matthews AG, Kuo AJ, Ramon-Maiques S, Han S, Champagne KS, Ivanov D, et al. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature. 2007 Dec 13;450(7172):1106-10.

McMurry MT, Krangel MS. A role for histone acetylation in the developmental regulation of VDJ recombination. Science. 2000 Jan 21;287(5452):495-8.

Meyer KD, Lin SC, Bernecky C, Gao Y, Taatjes DJ. p53 activates transcription by directing structural shifts in Mediator. Nat Struct Mol Biol. 2010 Jun;17(6):753-60.

Mills KD, Ferguson DO, Alt FW. The role of DNA breaks in genomic instability and tumorigenesis. - Immunological Reviews. 2003(- 1):- 77.

Mijuskovic M, Chou YF, Gigi V, Lindsay CR, Shestova O, Lewis SM, et al. Off-Target V(D)J Recombination Drives Lymphomagenesis and Is Escalated by Loss of the Rag2 C Terminus. Cell Rep. 2015 Sep 22;12(11):1842-52.

Morshead KB, Ciccone DN, Taverna SD, Allis CD, Oettinger MA. Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1 and flanked by peaks

of histone H3 dimethylated at lysine 4. Proc Natl Acad Sci U S A. 2003 Sep 30;100(20):11577-82.

Mostoslavsky R, Singh N, Kirillov A, Pelanda R, Cedar H, Chess A, et al. Kappa chain monoallelic demethylation and the establishment of allelic exclusion. Genes Dev. 1998 Jun 15;12(12):1801-11.

Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. Nature. 2014 Apr 17;508(7496):331-9.

Osipovich OA, Subrahmanyam R, Pierce S, Sen R, Oltz EM. Cutting edge: SWI/SNF mediates antisense Igh transcription and locus-wide accessibility in B cell precursors. J Immunol. 2009 Aug 1;183(3):1509-13.

Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, Tubio J, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. Nat Genet. 2014 Feb;46(2):116-25.

Perkins EJ, Kee BL, Ramsden DA. Histone 3 lysine 4 methylation during the pre-B to immature B-cell transition. Nucleic Acids Res. 2004 Mar 29;32(6):1942-7.

Perlot T, Alt FW, Bassing CH, Suh H, Pinaud E. Elucidation of IgH intronic enhancer functions via germ-line deletion. Proc Natl Acad Sci U S A. 2005 Oct 4;102(40):14362-7.

Puget N, Hirasawa R, Hu NS, Laviolette-Malirat N, Feil R, Khamlichi AA. Insertion of an imprinted insulator into the IgH locus reveals developmentally regulated, transcription-dependent control of V(D)J recombination. Mol Cell Biol. 2015 Feb;35(3):529-43.

Ramon-Maiques S, Kuo AJ, Carney D, Matthews AG, Oettinger MA, Gozani O, et al. The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. Proc Natl Acad Sci U S A. 2007 Nov 27;104(48):18993-8.

Ross AE, Vuica M, Desiderio S. Overlapping signals for protein degradation and nuclear localization define a role for intrinsic RAG-2 nuclear uptake in dividing cells. Mol Cell Biol. 2003 Aug;23(15):5308-19.

Roth DB, Roth SY. Unequal access: regulating V(D)J recombination through chromatin remodeling. Cell. 2000 Nov 22;103(5):699-702.

Ru H, Chambers MG, Fu TM, Tong AB, Liao M, Wu H. Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. Cell. 2015 Nov 19;163(5):1138-52.

Sadofsky MJ, Hesse JE, Gellert M. Definition of a core region of RAG-2 that is functional in V(D)J recombination. Nucleic Acids Res. 1994 May 25;22(10):1805-9.

Schatz DG, Ji Y. "Recombination Centres and the Orchestration of V(D)J Recombination." Nature Reviews Immunology 11.4 (2011): 251-63.

Schatz DG, Swanson PC. V(D)J recombination: mechanisms of initiation. Annu Rev Genet. 2011 March 11;45:167-202.

Schlissel MS, Corcoran LM, Baltimore D. Virus-transformed pre-B cells show ordered activation but not inactivation of immunoglobulin gene rearrangement and transcription. J Exp Med. 1991 Mar 1;173(3):711-20.

Sekiguchi JA, Whitlow S, Alt FW. Increased accumulation of hybrid V(D)J joins in cells expressing truncated versus full-length RAGs. Mol Cell. 2001 Dec;8(6):1383-90.

Selimyan R, Gerstein RM, Ivanova I, Precht P, Subrahmanyam R, Perlot T, et al. Localized DNA demethylation at recombination intermediates during immunoglobulin heavy chain gene assembly. PLoS Biol. 2013;11(1):e1001475.

Shetty K, Schatz DG. Recruitment of RAG1 and RAG2 to Chromatinized DNA during V(D)J Recombination. Mol Cell Biol. 2015 Nov;35(21):3701-13.

Shimazaki N, Askary A, Swanson PC, Lieber MR. Mechanistic basis for RAG discrimination between recombination sites and the off-target sites of human lymphomas. Mol Cell Biol. 2012 Jan;32(2):365-75.

Shimazaki N, Tsai AG, Lieber MR. H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: implications for translocations. Mol Cell. 2009 Jun 12;34(5):535-44.

Shimazaki N, Lieber MR. Histone methylation and V(D)J recombination. Int J Hematol. 2014 Sep;100(3):230-7.

Steen SB, Han JO, Mundy C, Oettinger MA, Roth DB. Roles of the "dispensable" portions of RAG-1 and RAG-2 in V(D)J recombination. Mol Cell Biol. 1999 Apr;19(4):3010-7.

Storb U, Arp B. Methylation patterns of immunoglobulin genes in lymphoid cells: correlation of expression and differentiation with undermethylation. Proc Natl Acad Sci U S A. 1983 Nov;80(21):6642-6.

Subrahmanyam R, Du H, Ivanova I, Chakraborty T, Ji Y, Zhang Y, et al. Localized epigenetic changes induced by DH recombination restricts recombinase to DJH junctions. Nat Immunol. 2012 Dec;13(12):1205-12.

Subrahmanyam R, Sen R. RAGs' eye view of the immunoglobulin heavy chain gene locus. Semin Immunol. 2010 Dec;22(6):337-45.

Sun Y, Jiang X, Xu Y, Ayrapetov MK, Moreau LA, Whetstine JR, et al. Histone H3 methylation links DNA damage detection to activation of the tumour suppressor Tip60. Nat Cell Biol. 2009 Nov;11(11):1376-82.

Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008 Jun;9(6):465-76.

Talukder SR, Dudley DD, Alt FW, Takahama Y, Akamatsu Y. Increased frequency of aberrant V(D)J recombination products in core RAG-expressing mice. Nucleic Acids Res. 2004 Aug 24;32(15):4539-49.

Van Ness BG, Weigert M, Coleclough C, Mather EL, Kelley DE, Perry RP. Transcription of the unrearranged mouse C kappa locus: sequence of the initiation region and comparison of activity with a rearranged V kappa-C kappa gene. Cell. 1981 Dec;27(3 Pt 2):593-602.

Ward A, Kumari G, Sen R, Desiderio S (submitted 2017 - Immunity) The RAG-2 inhibitory domain gates accessibility of the V(D)J recombinase to chromatin.

Yancopoulos GD, Alt FW. Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. Cell. 1985. 40: 271-281. J Immunol. 2012 Jan 1;188(1):10-20.

Yin FF, Bailey S, Innis CA, Ciubotaru M, Kamtekar S, Steitz TA, et al. Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. Nat Struct Mol Biol. 2009 May;16(5):499-508.

Yurchenko V, Xue Z, Sadofsky M. The RAG1 N-terminal domain is an E3 ubiquitin ligase. Genes Dev. 2003 Mar 1;17(5):581-5.

Zhang L, Reynolds TL, Shan X, Desiderio S. Coupling of V(D)J recombination to the cell cycle suppresses genomic instability and lymphoid tumorigenesis. Immunity. 2011 Feb 25;34(2):163-74.

Curriculum Vitae

The Johns Hopkins University School of Medicine

John Thomas Bettridge September 27th, 2017

Education:

Ph.D.	2017	Biochemistry, Cellular, and Molecular Biology, Hopkins University School of Medicine Advisor: Stephen Desiderio, M.D., Ph.D.
B.A.	2009	Biological Chemistry, *magnum cum laude* Religious Studies, c*um laude* Lawrence University, Appleton WI

Professional Experience:

Graduate Intern	2017	Papgene, Inc. *Authored an SBIR grant update for a new technology the company is developing.*
Research Assistant	2011-2017	Laboratory of Dr. Stephen Desiderio Johns Hopkins University School of Medicine *Demonstrated that an epigenetic post-translational mark, H3K4me3, is required for the generation of antigen receptor genes by developing lymphocytes. This epigenetic mark is an activating signal for the V(D)J recombinase.*
Research Specialist	2010-2011	Laboratory of Dr. Yoshihiro Kawaoka University of Wisconsin-Madison *Investigated molecular pathways of H1N1 influenza pathogenesis and infection using mouse and ferret animal models.*
Student Researcher	2008-2010	Laboratory of Dr. David Hall Lawrence University *Designed and created folic acid coated nanoparticles using CCMV virus capsid as a scaffold, for drug and gene delivery into cancer cells.*

Publications:

Bettridge J, Na CH, Pandey A, Desiderio S. H3K4me3 induces allosteric conformational changes in the DNA-binding and catalytic regions of the V(D)J recombinase. *Proceedings of the National Academy of Sciences*. 2017. PMC5338453.

Lu C, Ward A, Bettridge J, Liu Y, Desiderio S. An autoregulatory mechanism imposes allosteric control on the V(D)J recombinase by histone H3 methylation. *Cell Rep*. 2015; 10(1):29-38. PMC4308961.

Awards:

2012	Outstanding Presenter Award in Molecular Biology and Genetics
2005-2010	Lawrence University Alumni Academic Merit Scholarship
2010	Academic All-Conference Athlete, Swimming

Teaching Experience:

Teaching Assistant	2014-2015	Johns Hopkins University Genetics course Assistant to Dr. Jeremy Nathans and Dr. Roger Reeves. Guided review sessions and graded exams.
Teaching Assistant	2010	Lawrence University Genetics course Assistant to Dr. Elizabeth Destasio. ran the laboratory portion of the course, lectured, and graded exams.